Analsis dan Implementasi Algoritma *Winnowing* dengan *Synonym Recognition* pada Deteksi Plagiarisme untuk Dokumen Teks Berbahasa Indonesia

Jody[1], Agung Toto Wibowo[2], Anditya Arifianto[3]

[1.2.3] *Teknik Informatika, Fakultas Teknik Informatika, Telkom University*

Jalan Telekomunikasi No.1, Dayeuh Kolot, Bandung 40257
jody.muhammad.91@gmail.com[1], atwbox@gmail.com[2], anditya@telkomuniersity.ac.id[3]

Abstrak

Plagiarisme merupakan tindakan mengakui hasil karya orang lain sebagai hasil karya pribadi tanpa izin pemilik asli karya tersebut. Tindakan plagiarisme dalam bentuk dokumen sudah sangat banyak terjadi saat era digital seperti ini. Menyikapi hal tersebut, melalui Tugas Akhir ini penulis membangun sebuah sistem yang dapat digunakan untuk mendeteksi plagiarisme antar dokumen teks yaitu *Algoritma Winnowing* dengan *Synonym Recognition.*

Algoritma *Winnowing* adalah salah satu metode *document fingerprinting* yang digunakan untuk mendeteksi kemiripan antar teks dokumen dengan menggunakan teknik *hashing*. Algoritma ini dipilih karena *Winnowing* merupakan salah satu algoritma terbaik untuk mendapatkan nilai *similarity* antar teks dokumen baik dalam segi akurasi ataupum performansi. Dalam tugas akhir ini, algoritma *Winnowing* akan dikombinasikan dengan algoritma *Synonym Recognition* untuk mengatasi adanya parafrase dokumen yang dilakukan para pelaku plagiarisme.

Kata Kunci: Plagiarisme, algoritma *Winnowing, document fingerprinting, Similarity, Synonym Recognition*.

Abstract

Plagiarism is the act of acknowledging the work of other people personal work without the permission of the original owner of the work. The act of document plagiarism is prevalent in this digital era. In response to these problem, the purpose of this final project is to build a system that can be used to detect plagiarism between text documents that is Winnowing algorithm with Synonym Recognition.

Winnowing algorithm is one document fingerprinting method used to detect the similarity between text documents using hashing technique. Winnowing algorithm is chosen because it is one of the best algorithm to obtain the similarity value between the text of the document both in terms of accuracy or performance. In this final project, Winnowing algorithm will be combined with Synonym Recognition algorithms to deal with document paraphrasing committed by the perpetrators.

Keywords: Plagiarisme, Winnowing algorithm, document fingerprinting, Similarity, Synonym Recognition.

1. Pendahuluan

Plagiarisme adalah penjiplakan atau pengakuan atas karya orang lain oleh seseorang yang menjadikan karya tersebut sebagai karya ciptaannya. Orang yang melakukan plagiarisme disebut plagiaris atau plagiator. Dengan batasan demikian, plagiarisme adalah pencurian (bahasa kasarnya, pembajakan) dan plagiaris adalah pencuri (pembajak).

Plagiarisme tampaknya sudah membudaya, mudah terjadi dan sangat terasa pada jenjang pendidikan S1. Ketidaketatan kontrol bersumber dari kualitas pembimbing yang relatif rendah, jumlah pembimbing yang tidak proporsional dengan jumlah mahasiswa, dedikasi perguruan tinggi (PT) yang masih berorientasi pada keuntungan ekonomi semata, sikap tidak konsisten PT bersangkutan, sumber pustaka yang tidak mutakhir, akses ke sumber informasi yang sulit terjangkau, serta sanksi yang tidak jelas atau bahkan tidak ada terhadap plagiaris [14]. Contoh kasus yang terjadi belakangan ini adalah sebanyak 100 dosen setingkat guru besar, dosen dan lector kepala perguruan tinggi, melakukan plagiarisme pada 2012 lalu. Selama tahun 2012 kemarin ada empat yang diturunkan pangkatnya dan dua yang dipecat. Sesuai dengan Permendiknas No 17/2010 tentang Pencegahan dan Penanggulangan Plagiarisme di Pendidikan Tinggi, maka rektor harus menindak dosen yang melakukan plagiarisme tersebut. Selain itu, Kemendikbud juga menemukan sekitar 400 perguruan tinggi swasta, melakukan kejahatan berupa pemalsuan data jumlah dosen dan mahasiswa. Tindak pemalsuan tersebut dilakukan untuk mendapatkan dana pembinaan dan tunjangan sertifikasi dosen [15]. Menyadari hal itu, melalui tugas akhir ini penulis mengajukan sistem pendeteksi plagiarisme untuk memeriksa tingkat kemiripan antar dokumen.

Terdapat beberapa algoritma yang berfungsi untuk mendeteksi plagiarisme pada dokumen seperti *Rabin Karp* dan *Winnowing*. Pada proposal tugas akhir ini penulis memilih untuk menggunakan algoritma *Winnowing*. Algoritma *Winnowing* merupakan salah satu algoritma *Document Fingerprinting* yang menggunakan teknik *hashing* dalam mencocokkan dua atau lebih dokumen. Teknik *hashing* sendiri berguna untuk mengubah tiap string pada dokumen menjadi nilai *hash* yang nilainya akan digunakan sebagai *fingerprint* dari dokumen tersebut. Dalam *Winnowing* fungsi *hashing* yang digunakan adalah *Rolling Hashing*.

Algoritma *Winnowing* hanya dapat melihat kesamaan dalam karakter saja, sehingga apabila terjadi perubahan kata dengan makna yang sama, maka sistem tidak dapat mendeteksi kesamaan arti [5]. Oleh karena itu diperlukan suatu metode untuk mengatasi masalah tersebut, dan penulis menambahkan metode *Synonym Recognition*. *Synonim Recognition* adalah proses dilakukannya pencarian sinonim dari tiap kata, dan dilakukan pengubahan kata apabila terdapat sinonim dari kata yang dicari [9].

2. Landasan Teori

2.1 Plagiarisme

Dalam Kamus Bahasa Indonesia Kontemporer, Plagiarisme didefinisikan sebagai perbuatan menjiplak tulisan, ide dan sebagainya milik orang lain (KBIK, Hal. 1172). Sedangkan pada Kamus Besar Bahasa Indonesia, Plagiarisme diartikan sebagai penjiplakan yang melanggar hak cipta. Dengan demikian Plagiarisme adalah tindakan penyalahgunaan, pencurian/perampasan, penertiban, pernyataan, atau menyatakan sebagai milik sendiri dari sebuah pikiran, ide, tulisan, atau ciptaan yang sebenarnya milik orang lain[1].

Pada umumnya, sistem pendeteksi plagiarisme dikembangkan untuk:[2]

1. Data teks seperti essay, artikel, jurnal, penelitian dan sebagainya.
2. Dokumen teks yang lebih terstruktur seperti bahasa pemrograman.

Beberapa tipe plagiarisme antara lain sebagai berikut :[3]

1. *Word-for-word plagiarism*
   Tipe ini termasuk dalam tindakan menyalin setiap kata secara langsung tanpa diubah sedikitpun susunan kalimatnya.
2. *Plagiarism of authorship*
   Tindakan yang berupa mengakui hasil karya orang lain sebagai hasil karya sendiri dengan cara

mencantumkan nama sendiri dan menggantikan nama pengarang yang sebenarnya.

3. *Plagiarism of ideas*
   Tindakan yang berupa mengakui hasil pemikiran atau ide orang lain.
4. *Plagiarism of sources*
   Jika seorang penulis menggunakan kutipan dari penulis lainnya tanpa mencantumkan nama pemilik asli sumbernya.

Berdasarkan cara yang digunakan, Gipp dan Meuschke (2011) mengkategorikan praktek plagiat sebagai berikut:[10]

a. *Copy & Paste Plagiarism*, menyalin setiap kata tanpa ada perubahan.
b. *Disguised Plagiarism*, tergolong ke dalam praktek menutupi bagian yang disalin, teridentifikasi ke dalam empat teknik yaitu *shake & paste*, *expensive plagiarism*, *contractive plagiarism*, dan *mosaic plagiarism.*
c. *Technical Disguise*, teknik meringkas untuk menyembunyikan konten plagiat dari deteksi otomatis dengan memanfaatkan kelemahan dari metode analisis teks dasar, misal dengan mengganti huruf dengan simbol huruf asing.
d. *Undue Paraphrasing*, sengaja menuliskan ulang pemikiran asing dengan pemilihan kata dan gaya plagiator dengan menyembunyikan sumber asli.
e. *Translated Plagiarism*, mengkonversi konten dari satu bahasa ke bahasa lain.
f. *Idea Plagiarism*, menggunakan ide asing tanpa menyatakan sumber asli.
g. *Self Plagiarism*, penggunaan sebagian atau keseluruhan tulisan pribadi yang tidak dibenarkan secara ilmiah.

2.2 Winnowing

Algoritma *Winnowing* merupakan algoritma dokumen *fingerprinting* yang digunakan untuk mendeteksi salinan dokumen dengan menggunakan teknik *hashing*[2]. Menurut Cornic (2008), Algoritma *Winnowing* adalah metode yang meningkatkan efisiensi dari proses perbandingan dokumen *fingerprinting*. Dalam hal deteksi plagiarisme, metode ini dapat mengidentifikasi bagian-bagian kecil yang mirip dalam dokumen yang berjumlah banyak.

Input dari algoritma ini adalah dokumen teks yang diproses sehingga menghasilkan *output* berupa kumpulan nilai-nilai *hash*. Nilai *hash* merupakan nilai numerik yang dibentuk dari perhitungan ASCII setiap karakter. Kumpulan-kumpulan nilai *hash* tersebut selanjutnya disebut sebagai *fingerprint*. *Fingerprint* inilah yang digunakan sebagai indikator untuk membandingkan kemiripan antar dokumen teks [5].

Secara garis besar, konsep algoritma Winnowing bekerja adalah sebagai berikut[7]:

Gambar 2.1 Proses *Winnowing*

2.3 Langkah-langkah *Winnowing*

Diberikan contoh teks sebagai berikut:

.

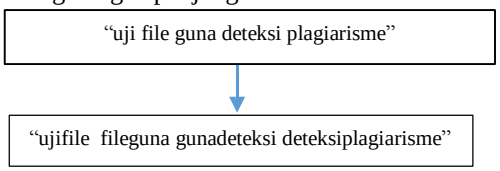
"Uji file guna deteksi plagiarisme"

1. Teks yang akan dideteksi yaitu Deteksi Plagiarisme. Langkah awal dalam penerapan algoritma *Winnowing* adalah melakukan *lowercase* atau mengubah tiap karakter pada string menjadi huruf kecil dan membuang karakter-karakter dari dokumen yang tidak relevan misal tanda baca, spasi dan simbol lain. Tanda baca, spasi dan simbol lain selain alfabet dikatakan tidak relevan karena tidak dapat diambil nilai uniknya dan tidak berkaitan dengan string yang akan diproses sehingga diperoleh hasil teks berikut:

"uji file guna deteksi plagiarisme"

2. Langkah kedua, membentuk teks hasil langkah pertama ke dalam rangkaian *k-gram*. Pada langkah ini, hasil teks dari langkah pertama yang berupa kumpulan string akan dikelompokkan menjadi kumpulan string yang baru dimana kumpulan string yang baru tersebut merupakan

hasil penggabungan string langkah pertama dengan panjang string yang digabungkan adalah *k*. Dibawah ini adalah contoh penggabungan string dengan panjang *k = 2* :



3. Langkah ketiga melakukan proses *Rolling Hash* untuk menghasilkan nilai *hash* dari setiap *gram* yang terbentuk. Contoh pada string "ujifile" dan nilai basis *hash* = 3, dengan menggunakan rumus

$$H(c_1..c_l) = c_1 \cdot b^{l-1} + c_2 \cdot b^{l-2} + \cdots + c_{l-1} \cdot b + c_l$$

Dimana:

- $H(c_1..c_l)$ = nilai hash
- c_l = nilai ASCII karakter ke-*l* pada string
- *l* = panjang string
- *b* = nilai basis *hash*

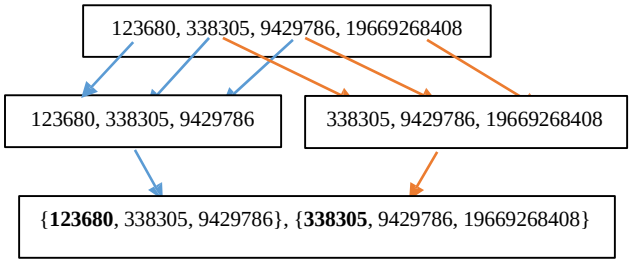
Maka

$$\begin{aligned} H(\text{"ujifile"}) = {} & (117 \text{ x } 3^{(7-1)}) + (106 \text{ x } 3^{(7-2)}) \\ & + (105 \text{ x } 3^{(7-3)}) + (102 \text{ x } 3^{(7-4)}) \\ & + (105 \text{ x } 3^{(7-5)}) + (108 \text{ x } 3^{(7-6)}) \\ & + (101) \\ & = 85293 + 25758 + 8505 + 2754 \\ & + 945 + 324 + 101 = 123680 \end{aligned}$$

Dari proses tersebut didapatkan nilai *hash* dari setiap *gram* sebagai berikut:

123680, 338305, 9429786, 19669268408

4. Setelah didapatkan nilai *hash* dari setiap *gram*, langkah keempat adalah membentuk *window*. Proses pembentukan *window* sama seperti proses *k-gram* dari nilai-nilai *hash* yang dihasilkan dengan besar *window* = 3:

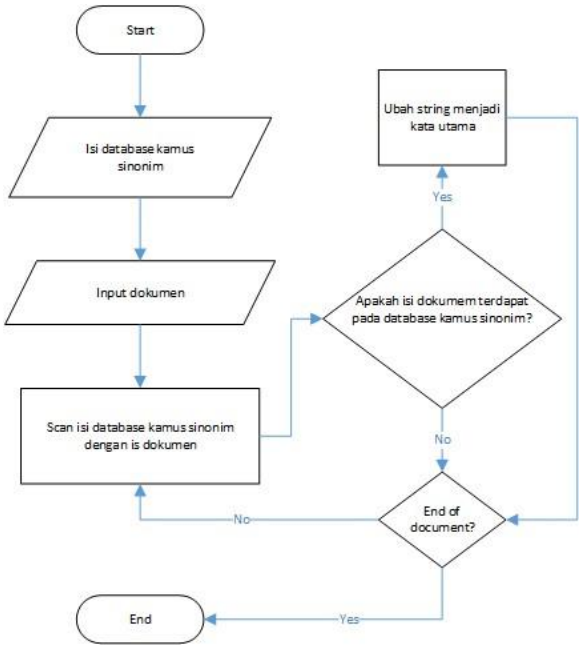


5. Langkah kelima adalah memilih nilai hash terkecil dari setiap window untuk dijadikan fingerprint dokumen tersebut. Dari window di atas didapatkan fingerprints berikut:

123680, 338305

2.4. *Synonym Recognition*

Synonym Recognition atau pengenalan sinonim adalah pendeteksian plagiarisme melalui pendekatan sinonim. Dalam hal ini, dokumen satu dibandingkan dengan dokumen lain dengan mendeteksi kata-kata yang mengandung sinonim. Dengan mendeteksi kata-kata yang memiliki kesamaan makna (sinonim) antara dokumen yang satu dengan yang lain dapat menambah nilai Similarity sehingga hasil deteksi plagiarisme lebih akurat. Synonym Recognition sangat bergantung pada database yang berisi kamus sinonim.[7]



Gambar 2.2 Proses *Synonym Recognition*

Berdasarkan flowchart di atas, secara garis besar proses *Synonym Recognition* adalah proses dimana seluruh isi dokumen yang dideteksi akan discan dan dicocokkan dengan kata yang ada dalam kamus sinonim kemudian akan diubah berdasarkan isi kamus sinonim tersebut. Jika *string* teks yang dicocokkan memiliki sinonim, kata tersebut akan dimasukkan ke dalam proses *Winnowing* dan dicocokkan kembali. Dan apabila *string* teks tidak memiliki sinonim, proses *Synonym Recognition* tidak

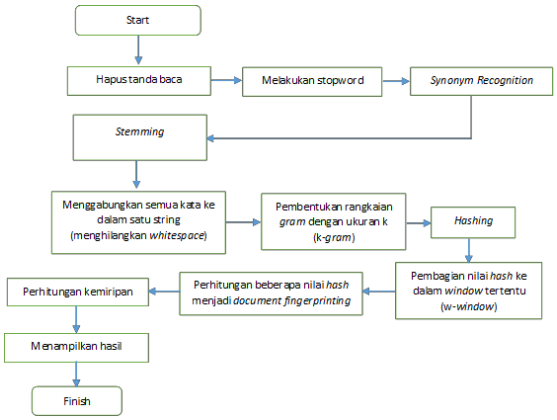
akan dilakukan. Langkah ini akan terus berulang sampai seluruh proses pencocokan *string* selesai[9]. Hubungan antara *Synonym Recognition* dengan *Winnowing* sangatlah penting. Algoritma *Winnowing* saja tidak dapat mendeteksi plagiarisme apabila terjadi perubahan kata yang memiliki arti yang sama. Banyaknya kasus plagiarisme yang hanya mengubah beberapa kata saja namun dengan arti yang sama mengakibatkan sulitnya algoritma *Winnowing* dalam mendeteksi kemiripan pada dokumen sehingga membutuhkan suatu algoritma yang mengatasi perubahan kata tersebut yaitu *Synonym Recognition*. *Synonym Recognition* akan memproses tiap kata pada dokumen dan mencari kata utama dari tiap kata tersebut ke dalam database. Setiap kata yang memiliki sinonim akan diubah menjadi kata utama berdasarkan database kamus sinonim sehingga secara keseleuruhan *Synonym Recognition* akan mengubah semua kata yang dianggap sebagai sinonim dari sebuah kata utama menjadi kata utama dan men-generalisasikannya menjadi kata utama.

Dokumen I : "uji file guna deteksi plagiarisme"
Dokumen II : "uji file guna deteksi penjiplakan"
Dengan Synonym Recognition maka kata plagiarisme akan diubah menjadi kata utamanya, yaitu penjiplakan, sehingga akan menjadi :
Dokumen I : "uji file guna deteksi penjiplakan"
Dokumen II : "uji file guna deteksi penjiplakan "

3. Perancangan Sistem

3.1 Analisa Kebutuhan Sistem

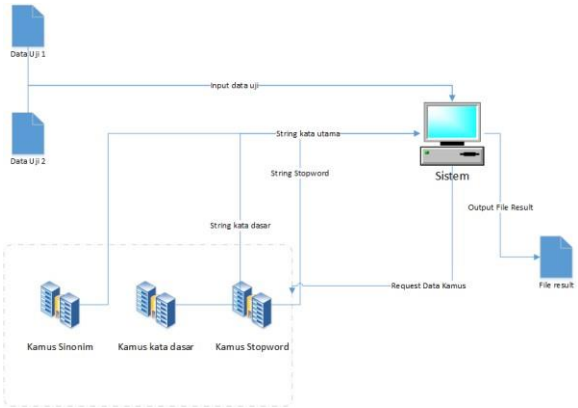
Dalam pembuatan tugas akhir ini, tahapan proses mengikuti proses *Winnowing* dengan ditambahkan proses *Synonym Recognition* sebelum masuk ke proses *Winnowing*. Tahapan proses dapat dilihat pada gambar 3.1.



Gambar 3.1 Tahapan Penelitian

3.2 Arsitektur Sistem

Dalam pembuatan tugas akhir ini, arsitektur sistem dapat dilihat pada gambar 3.2



Gambar 3.2 Arsitektur sistem

4. Hasil dan Pembahasan

4.1 Hasil Pengujian Parameter

Tabel 4.1 Hasil Pengujian parameter

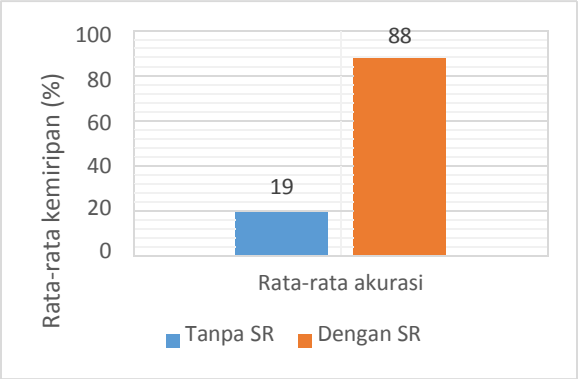
		Similarity (%)		Waktu (*s*)	
		Tanpa *SR*	**Dengan *SR***	**Tanpa *SR***	**Dengan *SR***
k	**1**	56,62	87,14	103,98	77,55
	2	39,19	77,16	89,23	70,64
	3	37,16	52,53	91,22	67,87
w	**4**	42,75	69,43	96,01	71,81
	5	42,47	73,46	89,87	67,21
	6	39,19	77,16	89,23	70,64
	7	41,28	79,73	88,8	71,17
	8	45,75	78,96	107,27	75,37
b	**5**	34,3	67,18	111,84	83,98
	6	45,3	78,45	96,29	68,56
	7	39,19	77,16	89,23	70,64
	8	35,11	67,18	87,61	77,87
	9	41,29	80,85	94,19	71

Pada tabel 4.1 diatas dapat dilihat pengaruh tiap parameter. Nilai *k-grams* berbanding terbalik dengan nilai *similarity* dan waktu. Semakin besar nilai *k-grams*, maka nilai *similarity* dan waktu akan semakin kecil. Sedangkan untuk nilai *w*, tidak terlalu berpengaruh terhadap nilai *similarity*. Tetapi dengan menggunakan *Synonym Recognition*, semakin besar nilai *w*, maka semakin besar nilai *similarity* yang didapat. Untuk pengaruh nilai *w* terhadap waktu yaitu semakin besar nilai *w*, maka semakin besar pula waktu yang diperlukan. Pengaruh nilai basis *hash* terhadap *similarity* yaitu semakin nilai *b* maka semakin besar nilai *similarity* walaupun perubahannya tidak terlalu signifikan. Untuk pengaruh nilai basis *hash* terhadap waktu yaitu semakin besar nilai *b*, maka waktu yang diperlukan semakin kecil. *Winnowing* yang dikombinasikan dengan *Synonym Recognition* menghasilkan rata-rata waktu yang lebih kecil dikarenakan pada tahap *Synonym Recognition*, beberapa kata yang memiliki sinonim diubah menjadi kata utama dimana kata utama tersebut sudah menjadi kata dasar. Hal ini menyebabkan kata tersebut tidak perlu dilakukan proses *stemming* dan waktu yang diperlukan menjadi lebih singkat.

Tabel 4.2 Jumlah kata yang melalui proses *stemming*

Dok	Jumlah kata		Perbedaan
	Tanpa *SR*	Dengan *SR*	
1	2475	1575	900
2	3069	2160	909
3	4022	2745	1277
4	1854	1681	173
5	1962	1296	666
6	4275	3258	1017
7	2664	2313	351
8	3987	2826	1161
9	7101	4284	2817
10	5013	3501	1512

4.2 Hasil Keakuratan Posisi Kalimat



Gambar 4.1 Hasil Akurasi Posisi Kalimat

Pada gambar 4.1 dapat dilihat bahwa penggunaan *Synonym Recognition* memiliki pengaruh yang sangat besar terhadap keakuratan posisi kalimat. Dari gambar 4.1 dapat dilihat perbedaan akurasi yang didapat mencapai 69%. Hal ini membuktikan bahwa penggunaan *Synonym Recognition* pada *Winnowing* sangat berguna dalam akurasi posisi kalimat.

4.3 Hasil Pengujian Kemiripan Dokumen

Tabel 4.3 Hasil pengujian kemiripan dokumen

No	Pengembangan Kalimat (%)		Parafrase Dokumen (%)	
	Tanpa *SR*	**Dengan *SR***	**Tanpa *SR***	**Dengan *SR***
1	18,68	29,3	56,67	82,32
2	12,22	31,58	63,6	74,94

3	14,58	43,1	59,66	64,66
4	14,65	22,69	50,41	65,92
5	16,52	34,99	47,91	69,99
6	18,49	30,78	47,15	57,55
7	6,35	31,84	52,01	58,18
8	8,86	25,8	59,15	65,75
9	23,19	40,15	69,94	72,28
10	12,34	28,75	58,02	72,28

Pada tabel 4.3 diatas dapat dilihat bahwa nilai *similarity* yang dihasilkan *Winnowing* dengan *Synonym Recognition* lebih besar dibandingkan tanpa *Synonym Recognition*. Hal ini berlaku untuk kedua jenis pengujian baik itu pengembangan kalimat ataupun parafrase dokumen. Rata-rata keseluruhan kenaikan nilai *similarity* yang terjadi sebesar 14,62%.

Tabel 4.4 Pengaruh *Synonym Recognition* terhadap nilai *similarity*

Jenis Pengujian	**Perubahan (%)**		
	Naik	**Turun**	**Tetap**
Pengembangan Kalimat	83,334	12,667	3,999
Parafrase Dokumen	83,113	15,777	1,11

Pada tabel 4.4 diatas dapat dilihat bahwa pengaruh *Synonym Recognition* pada *Winnowing* sebagian besar mengalami kenaikan nilai *similarity*. Kenaikan yang terjadi lebih dari 83%. Hal ini membuktikan bawa *Synonym Recognition* memiliki pengaruh yang cukup besar dalam mendapatkan nilai *similarity*. Selain kenaikan, pengaruh *Synonym Recognition* juga mengalami penurunan nilai *similarity*. Penurunan terjadi karena adanya kata ambigu pada data uji. Hal ini menyebabkan sistem tidak dapat mengganti kata sinonim secara tepat dan menyebabkan bertambahnya perbedaan antar dokumen. Oleh karena itu terjadinya penurunan nilai *similarity*.

Tabel 4.5 Persentase pengujian dengan nilai *similarity* lebih dari nilai *threshold*

No	**Tanpa *SR* (%)**	**Dengan *SR* (%)**
1	68,89	100
2	93,33	100
3	80	97,88
4	46,67	77,78
5	35,56	88,89
6	44,44	62,22
7	44,44	60
8	84,44	86,67
9	100	100
10	77,78	100

Pada tabel 4.5 dapat dilihat bahwa *Winnowing* yang ditambahkan dengan *Synonym Recognition*, menaikkan persentase keberhasilan pendeteksian kemiripan. Tingkat keberhasilan pendeteksian yaitu hasil pengujian yang memiliki nilai *similarity* lebih dari 50%. Dari seluruh pengujian yang dilakukan, rata-rata pengujian mengalami kenaikan jumlah perbandingan yang lulus *threshold* atau berhasil terdeteksi adanya kemiripan. Nilai terbesar bahkan mencapai 100% dari tipe dokumen pengujiannya. Hal ini terjadi karena *Synonym Recognition* dapat meningkatkan nilai *similarity* dari hasil *Winnowing* yang dimana nilai *similarity* dari *Winnowing* pada saat menangani kasus parafrase mendapatkan nilai *similarity* yang kecil. Sedangkan *Winnowing* yang dikombinasikan dengan *Synonym Recognition* dapat menangani kasus parafrase dan meningkatkan nilai *similarity*.

4.4 Hasil Perhitungan Akurasi

Tabel 4.6 Hasil perhitungan akurasi

Dok	**Tanpa *SR***	**Dengan *SR***
	Akurasi	**Akurasi**
1	77.78	91.11
2	68.89	80.00
3	60.00	64.44
4	84.44	91.11
5	80.00	88.89
6	75.56	84.44
7	64.44	88.89
8	82.22	88.89
9	66.67	88.89
10	80.00	86.67

Pada tabel 4.6 dapat dilihat bahwa nilai akurasi *Winnowing* yang dikombinasikan dengan *Synonym Recognition* mengalami kenaikan dibandingkan nilai akurasi *Winnowing* tanpa *Synonym Recognition*. Hal ini dikarenakan nilai *similarity* yang didapat dengan menambahkan *Synonym Recognition* lebih tinggi dibandingkan tanpa *Synonym Recognition*. Nilai

similarity tersebut sangat berpengaruh terhadap akurasi sistem dalam mendeteksi plagiarisme. Dokumen akan dikatakan hasil plagiarisme jika nilai *similarity* lebih dari 50. Dengan menambahkan *Synonym Recognition*, nilai *similarity* yang didapat menjadi meningkat melebihi nilai *threshold* yaitu 50 sehingga dokumen yang dibandingkan dikatakan plagiarisme. Hal ini menyebabkan nilai akurasi yang didapat menjadi meningkat.

5 Kesimpulan dan Saran

5.1 Kesimpulan

Kesimpulan yang bisa diambil dari proses penelitian tugas akhir ini adalah sebagai berikut:

1. Berdasarkan percobaan yang dilakukan pada, didapat bahwa parameter terbaik untuk *Winnowing* dengan *Synonym Recognition* pada kasus ini adalah *k-grams* kata = 1, *window* = 6 dan nilai basis *hash* = 7.

2. Berdasarkan percobaan yang dilakukan, didapat bahwa waktu yang diperlukan oleh *Winnowing* ditambahkan dengan *Synonym Recognition* lebih kecil dibandingkan dengan *Winnowing* tanpa *Synonym Recognition* dikarenakan pada tahap *Synonym Recognition*, beberapa kata yang memiliki sinonim diubah menjadi kata utama dimana kata utama tersebut sudah menjadi kata dasar. Hal ini menyebabkan kata tersebut tidak perlu dilakukan proses *stemming* dan waktu yang diperlukan menjadi lebih singkat.

3. Berdasarkan percobaan yang dilakukan dapat dilihat bahwa penggunaan *Synonym Recognition* pada *Winnowing* sangat berpengaruh dalam menentukan kalimat termirip dalam satu dokumen teks dengan nilai keberhasilan sebesar 88% dimana jauh berbeda dibandingkan dengan tanpa *Synonym Recognition* yang hanya menghasilkan nilai keberhasilan sebesar 19%.

4. Berdasarkan pecobaan yang dilakukan, dapat dilihat bahwa penggunaan *Synonym Recognition* pada *Winnowing* untuk kasus pengembangan kalimat memiliki pengaruh yang baik dalam mendeteksi kemiripan. Hal ini dapat dilihat dari meningkatnya nilai rata-rata kemiripan yang diperoleh oleh *Winnowing* dengan *Synonym Recognition* dibandingkan tanpa *Synonym Recognition* yaitu 14.58% menjadi 31.9%. Selain itu pula dapat dilihat jumlah persentase pengujian yang mengalami kenaikan mencapai 83.33% dari seluruh pengujian. Kedua hal ini membuktikan bahwa *Synonym Recognition* pada *Winnowing* memiliki pengaruh menaikkan tingkat keakuratan dalam mendeteksi kemiripan dokumen.

5. Berdasarkan percobaan yang dilakukan pada 4.4.3 poin b, dapat dilihat bahwa pengaruh *Synonym Recognition* pada *Winnowing* dalam mendeteksi kemiripan dokumen adalah meningkatkan keakuratan nilai kemiripan. Dari seluruh pengujian yang dilakukan, terjadi kenaikan rata-rata kemiripan dengan nilai rata-rata kenaikan sebesar 12.16%. Selain itu, jumlah perbandingan yang memiliki nilai kemiripan melebihi nilai threshold atau 50% mengalami kenaikan sebesar 19.78%. Semua hal tersebut membuktikan bahwa pengaruh *Synonym Recognition* pada *Winnowing* terhadap keakuratan mendeteksi kemiripan lebih baik dibandingkan *Winnowing* tanpa *Synonym Recognition*.

5.2 Saran

Beberapa saran yang diajukan untuk kelanjutan penelitian tugas akhir ini adalah sebagai berikut:

1. Kelemahan dari *Synonym Recognition* dalam pengujian ini adalah ambiguitas dari sebuah kata. Dimana sinonim dari kata yang ambigu sangat sulit dikenali. Oleh karena itu diharapkan dalam pengujian ke depannya adanya suatu rancangan yang dapat mengatasi kasus ambiguitas untuk mendapatkan hasil yang jauh lebih baik.

2. Untuk memudahkan sistem dalam mendeteksi tindakan plagiat dokumen teks, input untuk sistem diharapkan merupakan dokumen yang memiliki satu topik. Sehingga akan lebih mudah dalam mendeteksi plagiat tersebut.

3. Peran database kamus sinonim dalam pengujian ini sangatlah penting. Oleh karena itu semakin lengkap database yang dimiliki sistem, maka semakin besar kemungkinan menaiknya keakuratan sistem dalam mendeteksi kemiripan dokumen.

Daftar Pustaka

[1] *Aplikasi Pendeteksi Duplikasi Dokumen Teks Bahasa Indonesia Menggunakan Algoritma Winnowig Dengan Metode K-Gram Dan Synonym Recognition. Jurusan Teknik Informatika Universitas Muhammadiyah Malang.* **Pratama, R., Cahyono, B. and Marthasari, I.**

[2] *Deteksi Plagiarisme Dokumen Teks Menggunakan Algoritma Rabin-Karp dengan Synonym Recognition. Program Sudi Ilmu Komputer, Program Teknologi Informatika dan Ilmu Komputer, Universitas Brawijaya.* **Dewanto, S., Indriati and Cholissodin, I.**

[3] *Document Silimarity Analysis for a Plagiarism Detection System, 2nd Indian International Conference on Artificial Intelegence (IICAI-05), pp. 2534-2544.* **Iyer, Parvati and Abhipsita, Singh. 2005.** 2005.

[4] **2014.** Frasa. *http://id.wikipedia.org/wiki/Frasa*. [Online] September 18, 2014.

[5] **2014.** Lakukan plagiat, 100 dosen di Indonesia dapat sanksi. *http://nasional.sindonews.com/read/2013/10/02/15/790041/lakukan-plagiat-100-dosen-di-indonesia-dapat-sanksi.* [Online] May 5, 2014.

[6] *Near Similiarity Search and Plagiarism Analysis, 29th Annual Conference of the German Classification Society (GfK1), Magdeburg, ISDN 1431-8814, pp. 430-437.* **Stein, B. and Eissen, S. Meyer zu. 2006.** 2006.

[7] **2014.** Sørensen–Dice coefficient. *http://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient.* [Online] May 5, 2014.

[8] **Nugroho, Eko. 2011.** *Perancangan Sistem Deteksi Plagiarisme Dokumen Teks Dengan Menggunakan Algoritma Rabin-Karp. Program Studi Ilmu Komputer, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya.* 2011.

[9] *Perbandingan Pendekatan Deteksi Plagiarisme Dokumen Dalam Bahasa Inggris : Proceeding, Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2008), Auditorium Universitas Gunadarma, Depok, 20-21 Agustus 2008;ISSN : 1411 - 6286.* **Kurniawati, Ana and Wicaksana, I Wayan Simri. 2008.** Depok : s.n., 2008.

[10] **2014.** PLAGIARISME, KESALAHAN BERBAHASA TULIS, DAN PENANGGULANGANNYA. *http://www.dikti.go.id/id/2013/02/25/plagiarisme-kesalahan-berbahasa-tulis-dan-penanggulangannya/.* [Online] May 5, 2014.

[11] **Pratama, Indra. 2013.** *Analisis dan Implementasi Perbandingan Stemming Menggunakan Algoritma Ahmad Yusoff Sembok dengan Jelita Asian pada Pencarian Ayat Al-Quran yang Terkait Hadits.* Bandung: IT Telkom : s.n., 2013.

[12] **Redaksi, Tim. 2008.** *Tesaurus Bahasa Indonesia Pusat Bahasa. Pusat Bahasa, Departemen Pendidikan Nasional.* 2008.

[13] **Ridhatillah, Ardini. 2003.** *Dealing with Plagiarism in the Information System Research Community : A look at Factors that Drive Plagiarism and Ways to Addrees Them, MIS Quarterly;Vol. 27, No. 4, p. 511-532/December 2003.* 2003.

[14] **Salmuasih.** *Perancangan Sistem Deteksi Plagiat pada Dokumen Teks dengan Konsep Similarity menggunakan Algortima Rabin Karp. Sekolah Tinggi Manajemen Informatika dan Komputer, AMIKOM.* Yogyakarta : s.n.

[15] *Winnowing: Local Algorithms for Document Fingerprinting. In Proceedings of the ACMSIGMOD International Conference on Management of Data, June 9-12, San Diego, CA; ACM 1-58113-634-X/03/06. pp 76-85.* **Schleimer, S., Wilkerson, D. and Aiken, A. 2003**.

[16] *Implementasi Algoritma Winnowing untuk Mendeteksi Kemiripan pada Dokumen Teks,INFORMATIKA; Vol. 9, N0. 1, April 2013, pp 73-81.* **Kharisman, O., Susanto, B. and Suwarno, S. 2013.**

[17] **Asian, Jelita, Dkk. 2005.** *Stemming Indonesian. Australia : School of Computer Science and Information Technology*.

[18] **Nazief, B. A. A. & Adriani, M. 1996.** *Confixstripping: Approach to Stemming Algorithm for Bahasa Indonesia. Internal publication, Faculty of Science, University of Indonesia*.