

## BAB V

### PENGUJIAN DAN ANALISIS

#### 5.1. Pengujian Sistem

Bab ini menjelaskan pengujian sistem setelah diimplementasikan di bab sebelumnya dengan tujuan apakah sistem sudah berjalan dengan baik dan sesuai kebutuhan atau belum.

#### 5.2. Data Pengujian

Untuk melakukan uji coba sistem mengambil tiap judul jurnal dan membandingkan dengan data yang sudah dimanipulasi. Sistem mengambil *sample* seleksi fitur bagian pendahuluan dari tiap jurnal. Dari tiap sampel tersebut dilakukan perubahan sedemikian rupa hingga memiliki kesamaan kata sesuai yang dibutuhkan. Adapun dokumen teks yang diuji adalah sebagai berikut.

Tabel 5.1. Data uji dokumen teks.

No	Nama	Kesamaan (%)	Size (KB)
1	Deny-ori.docx	Dokumen asli	17
2	Deny-x1.docx	80	17
3	Deny-x2.docx	60	17
4	Deny-x3.docx	40	17
5	Deny-x4.docx	20	17
6	Hima-ori.docx	Dokumen asli	17
7	Hima-x1.docx	80	17
8	Hima-x1.docx	60	17
9	Hima-x1.docx	40	17
10	Hima-x1.docx	20	17

#### 5.3. Hasil Percobaan Sistem

##### 5.3.1. Hasil Uji Coba *K-gram*

Sistem membandingkan dokumen asli Deny-ori.docx dengan Deny-x1.docx dengan parameter *k-gram* yang diujicoba dari 1 hingga 5 dan modulo tetap. Sistem melihat dari hasil kesamaan kata dan waktu proses. Hasil pengujian tampak pada tabel 5.2 sampai 5.11.

Tabel 5.2. Tabel Uji *k-gram* dengan modulo 79

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	79	100	0,25
2	2	79	100	0,25
3	3	79	99,62	0,27
4	4	79	98,63	0,27
5	5	79	97,98	0,27

Tabel 5.3. Tabel Uji *k-gram* dengan modulo 83

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	83	100	0,25
2	2	83	100	0,25
3	3	83	99,62	0,27
4	4	83	98,63	0,27
5	5	83	97,98	0,52

Tabel 5.4. Tabel Uji *k-gram* dengan modulo 89

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	89	100	0,25
2	2	89	100	0,25
3	3	89	99,62	0,27
4	4	89	98,63	0,27
5	5	89	97,98	0,52

Tabel 5.5. Tabel Uji *k-gram* dengan modulo 97

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	97	100	0,3
2	2	97	100	0,35
3	3	97	99,62	0,3
4	4	97	98,63	0,32
5	5	97	97,98	0,52

Tabel 5.6. Tabel Uji *k-gram* dengan modulo 101

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	101	100	0,25
2	2	101	100	0,27
3	3	101	99,62	0,25
4	4	101	98,63	0,28
5	5	101	97,98	0,52

Tabel 5.7. Tabel Uji *k-gram* dengan modulo 103

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	103	100	0,3
2	2	103	100	0,33
3	3	103	99,62	0,25
4	4	103	98,63	0,47
5	5	103	97,98	0,8

Tabel 5.8. Tabel Uji *k-gram* dengan modulo 127

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	127	100	0,33
2	2	127	100	0,33
3	3	127	99,62	0,22
4	4	127	98,63	0,25
5	5	127	97,98	0,65

Tabel 5.9. Tabel Uji *k-gram* dengan modulo 151

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	151	100	0,35
2	2	151	100	0,45
3	3	151	99,62	0,4
4	4	151	98,63	0,2
5	5	151	97,98	0,52

Tabel 5.10. Tabel Uji *k-gram* dengan modulo 173

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	173	100	0,33
2	2	173	100	0,52
3	3	173	99,62	0,38
4	4	173	98,63	0,27
5	5	173	97,98	0,58

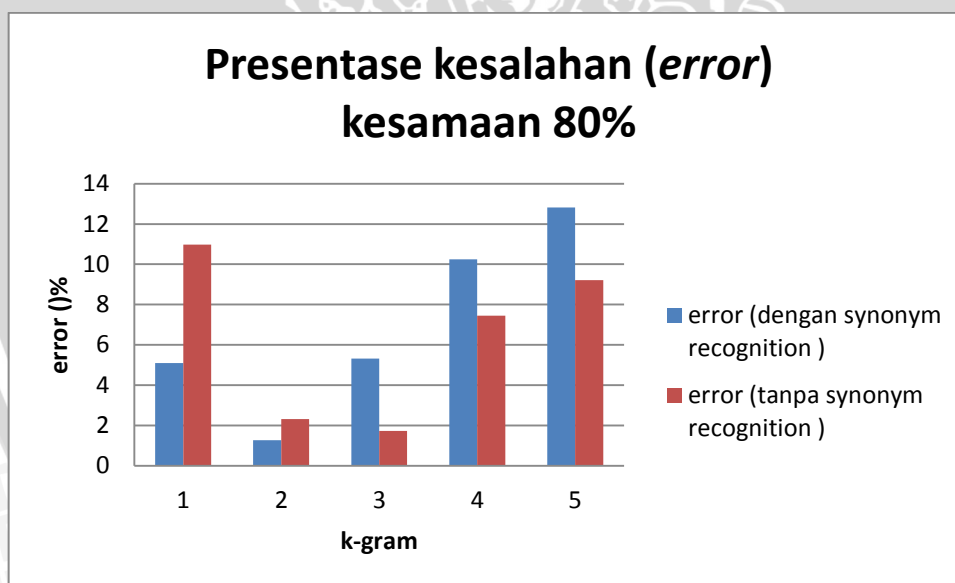
Tabel 5.11. Tabel Uji *k-gram* dengan modulo 251

No	<i>K-gram</i>	Modulo	Nilai Kesamaan (%)	Waktu Proses (detik)
1	1	251	100	0,43
2	2	251	100	0,37
3	3	251	99,62	0,4
4	4	251	98,63	0,43
5	5	251	97,98	0,68

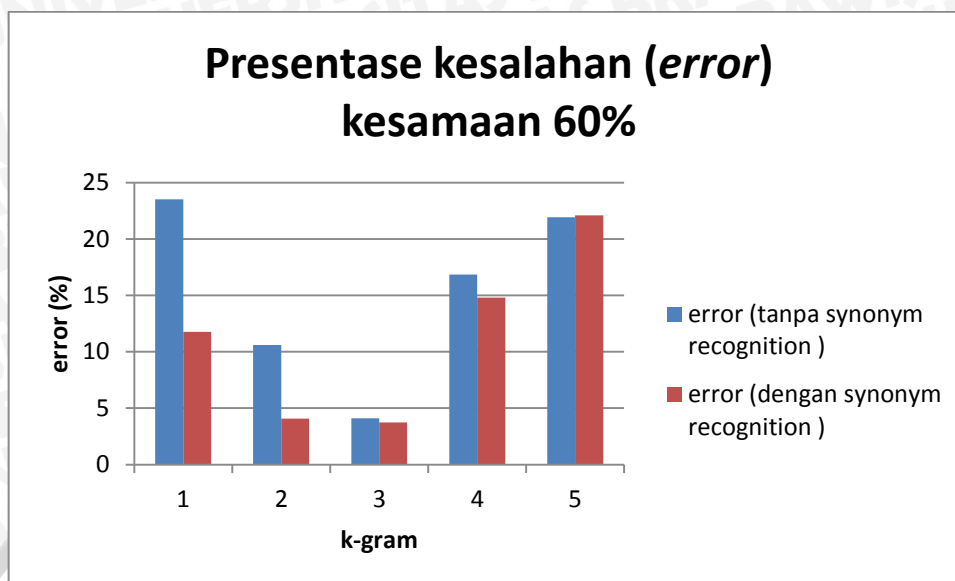


### 5.3.2. Hasil Uji Coba Presentase Kesalahan (*Error*)

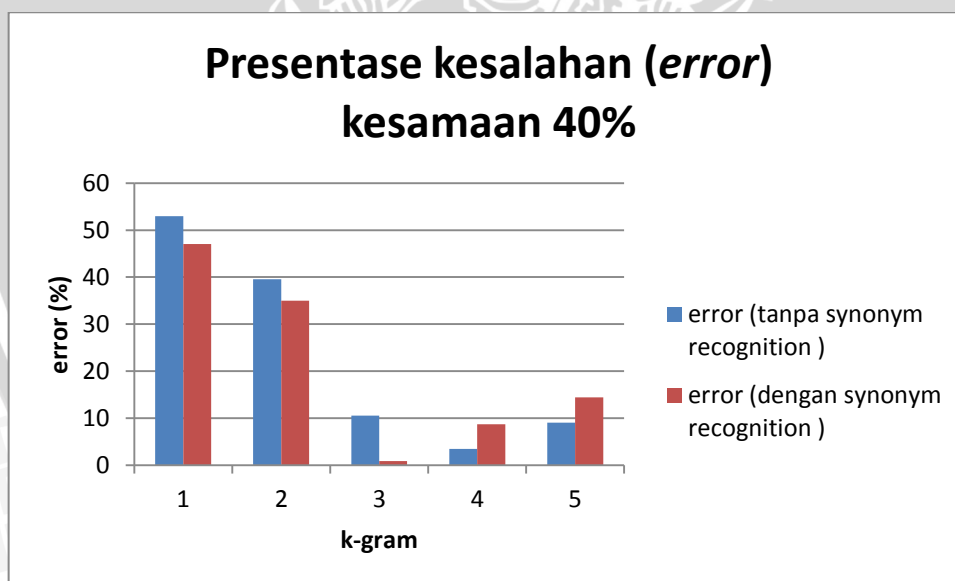
Uji coba ini membandingkan dokumen asli xDeny-ori.docx dengan masing-masing dokumen uji xDeny-x1.docx, xDeny-x2.docx, xDeny-x3.docx, dan xDeny-x4.docx. Keempat dokumen uji ini merupakan *copy*-an dari dokumen asli yang sudah dimodifikasi. Dengan memakai *k-gram* 1 sampai 5, dan modulo 101 dicari presentase kesalahan (*error*) dari proses pencocokan *string*. Presentase kesalahan di sini adalah selisih nilai kesamaan *substring* (hasil pengujian) dengan nilai kesamaan *string*. Nilai kesamaan *string* sendiri diperoleh dari jumlah *string* sama dibagi total *string* dokumen asli. Presentase kesalahan menyatakan perbedaan nilai kesamaan hasil pengujian dengan hasil perbandingan antar *string* yang sesungguhnya. Tabel hasil perhitungan presentase kesalahan ditampilkan dalam Lampiran II. Hasilnya bisa dilihat pada diagram seperti pada Gambar 5.1 sampai Gambar 5.4.



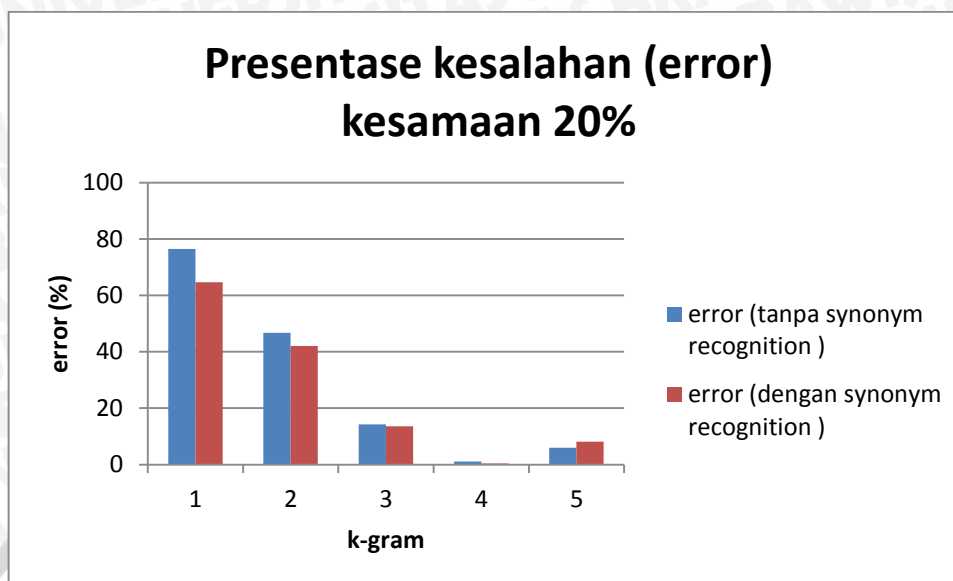
Gambar 5.1. Grafik perbandingan presentase kesalahan dokumen asli dengan xDeny-x1.docx yang memuat 80% kesamaan, dengan dan tanpa *synonym recognition*.



Gambar 5.2. Grafik perbandingan presentase kesalahan dokumen asli dengan xDeny-x2.docx yang memuat 60% kesamaan, dengan dan tanpa *synonym recognition*.



Gambar 5.3. Grafik perbandingan presentase kesalahan dokumen asli dengan xDeny-x3.docx yang memuat 40% kesamaan, dengan dan tanpa *synonym recognition*.



Gambar 5.4. Grafik perbandingan presentase kesalahan dokumen asli dengan xDeny-x4.docx yang memuat 20% kesamaan, dengan dan tanpa *synonym recognition*.

Grafik presentase kesalahan di atas menunjukkan bahwa kesalahan (*error*) paling besar terjadi pada perbandingan dokumen asli dengan xDeny-x4.docx yang memiliki 20% kesamaan. Dan *margin error* paling kecil terjadi pada perbandingan terhadap dokumen dengan kesamaan 80%.

Terlihat juga bahwa presentase kesalahan cenderung membentuk kurva dengan nilai *error* terkecil rata-rata terletak pada k-gram 3, 4, dan 2. Dan hampir di semua kasus presentase *error* sistem tanpa *synonym recognition* lebih besar daripada yang menggunakan *synonym recognition*.

### 5.3.3. Uji Coba *Synonym Recognition*

Pengujian terhadap metode *synonym recognition* bertujuan untuk mengetahui pengaruhnya terhadap tingkat kesamaan kata (%) dan waktu proses (detik). Pengujian dilakukan terhadap data uji Deny-ori.docx dan dokumen *copy*-nya yang sudah dimanipulasi. Sistem menggunakan *k-gram* 4 dan modulo 101. Proses pengujian dapat dilihat di tabel 5.15



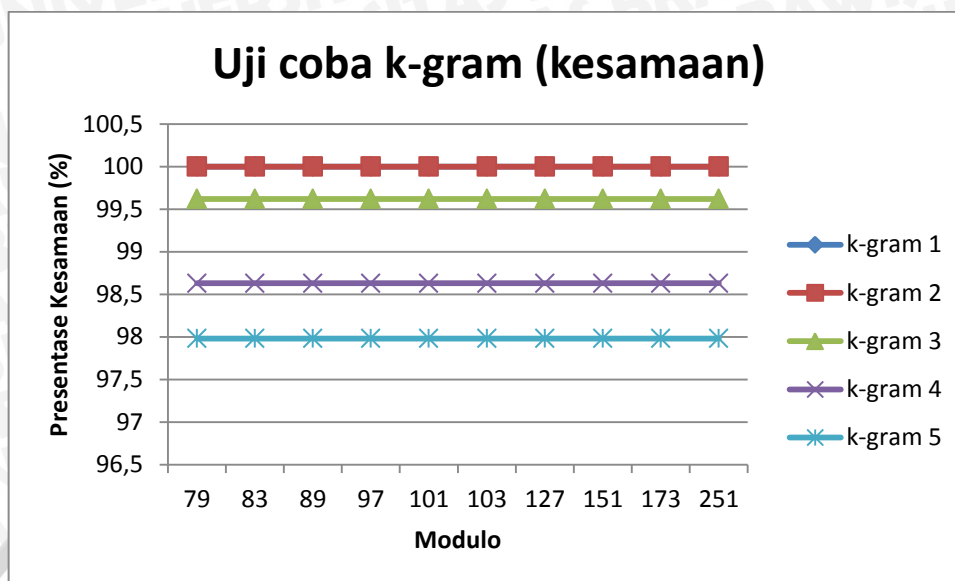
Tabel 5.12. Uji coba *synonym recognition*.

No	Dokumen Asli	Dokumen Uji	Synonym Recognition	Nilai Kesamaan (%)	Waktu Proses (detik)
1	Deny-ori.docx	Deny-ori.docx	Tidak	100	1,83
2		Deny-x1.docx	Tidak	94	1,83
3		Deny-x2.docx	Tidak	68,09	1,83
4		Deny-x3.docx	Tidak	43,11	1,83
5		Deny-x4.docx	Tidak	27,75	1,82
6	Deny-ori.docx	Deny-ori.docx	Ya	100	2,34
7		Deny-x1.docx	Ya	94,83	2,35
8		Deny-x2.docx	Ya	70,75	2,33
9		Deny-x3.docx	Ya	43,13	2,34
10		Deny-x4.docx	Ya	27,99	2,35
11	Hima-ori.docx	Hima-ori.docx	Tidak	100	1,82
12		Hima-x1.docx	Tidak	94,77	1,83
13		Hima-x2.docx	Tidak	63,42	1,82
14		Hima-x3.docx	Tidak	46,93	1,83
15		Hima-x4.docx	Tidak	26,13	1,83
16	Hima-ori.docx	Hima-ori.docx	Ya	100	2,4
17		Hima-x1.docx	Ya	94,97	2,3
18		Hima-x2.docx	Ya	68,84	2,3
19		Hima-x3.docx	Ya	46,93	2,4
20		Hima-x4.docx	Ya	27,03	2,3

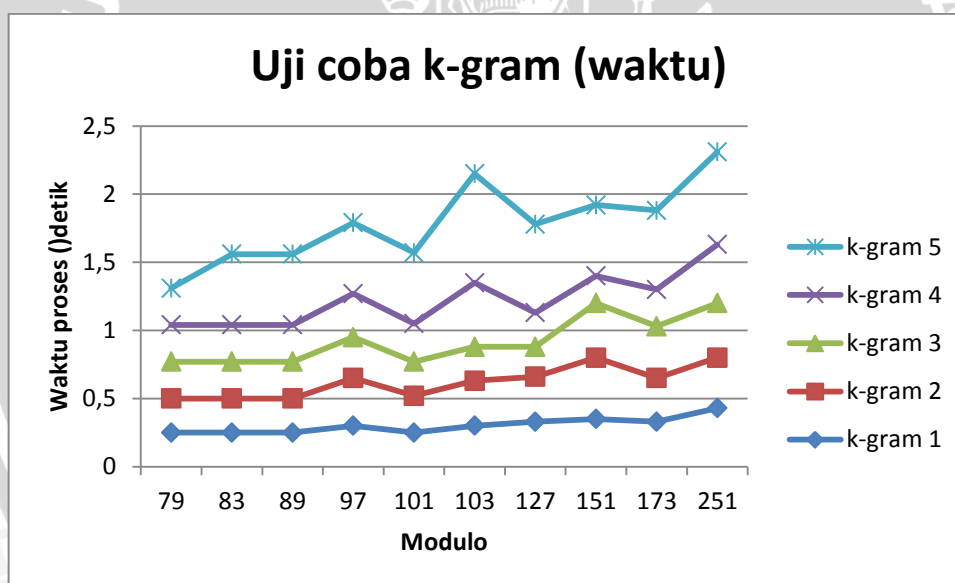
## 5.4. Analisa Hasil

### 5.4.1. Analisa Hasil Parameter *k-gram*

Berdasarkan hasil pengujian sebelumnya dari tabel 5.1 sampai tabel 5.10 terlihat bahwa *k-gram* tidak mempengaruhi nilai kesamaan untuk tiap modulo. Tingkat kesamaan kata berbanding terbalik dengan nilai *k-gram*. Semakin kecil nilai *k-gram*, semakin banyak string yang ditemukan cocok. Hal ini membuat tingkat kesamaan tinggi. Karena semakin pendek *substring k-gram* akan semakin mudah menemukan kesamaan *substring*. Sehingga untuk *k-gram* 1 setiap karakter tunggal di dokumen uji akan cocok dengan *substring* pada dokumen asli.



Gambar 5.5. Grafik perbandingan kesamaan kata untuk modulo 79 sampai 251.



Gambar 5.6. Grafik perbandingan waktu proses untuk modulo 79 sampai 251.

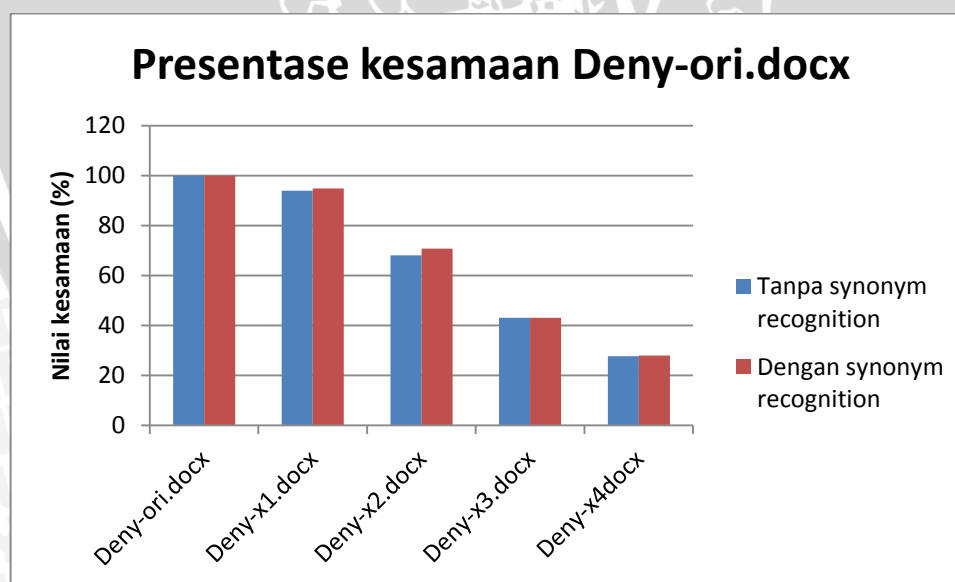
Pada Gambar 5.5 terlihat bahwa parameter *k-gram* tidak berpengaruh terhadap presentase kesamaan kata yang dihasilkan pada setiap modulo. Pada uji coba dengan melihat waktu proses *k-gram* ternyata mempengaruhi hasil yang dikeluarkan. Semakin tinggi nilai *k-gram* semakin banyak juga waktu proses yang dibutuhkan. Dan untuk modulo sendiri waktu proses cenderung meningkat berbanding lurus dengan nilai modulo meskipun garis pada grafik naik turun.



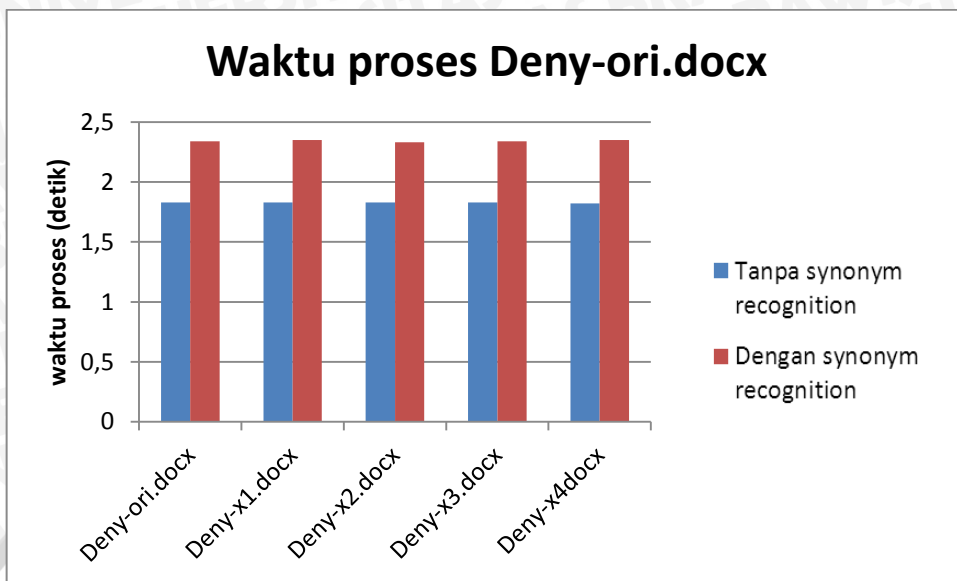
Dari analisa grafik di atas bisa disimpulkan jika semakin besar tingkat kesamaan tidak menjamin keakuratan kesamaan *string*. Meskipun nilai kesamaan yang dihasilkan pengujian cenderung lebih besar pada dokumen uji yang memiliki kesamaan 80%, keluaran untuk k-gram 1 tidak terlalu akurat. Terbukti dengan lebarnya jarak antara sistem yang memakai *synonym recognition* dengan yang tidak memakai *synonym recognition*. Dan waktu proses lebih banyak dipengaruhi oleh nilai *k-gram* dibanding modulo.

#### 5.4.2. Analisa Hasil *Synonym Recognition*

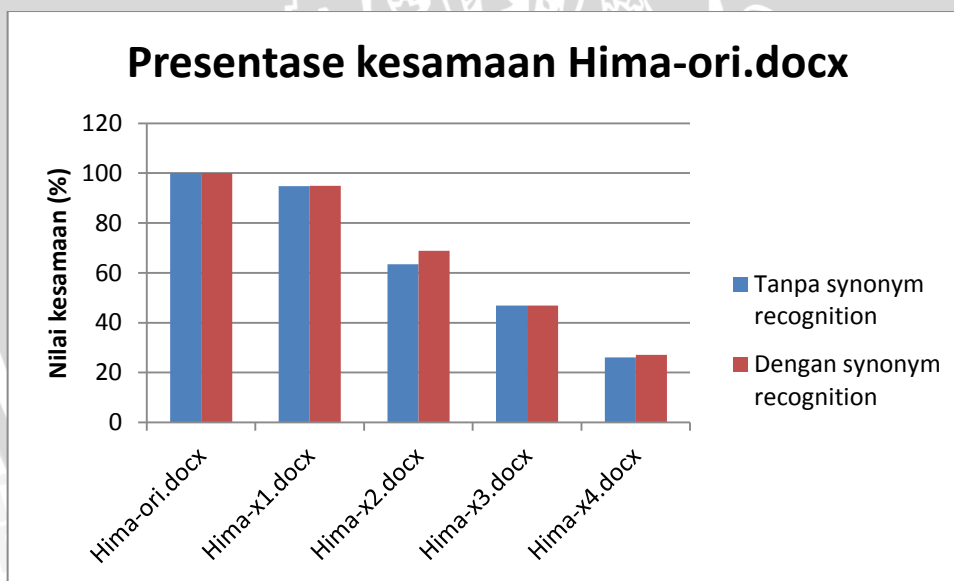
Untuk mengetahui seberapa besar pengaruh fitur *synonym recognition* terhadap sistem deteksi plagiarisme telah dilakukan beberapa uji coba menggunakan dua dokumen. Yang pertama menggunakan dokumen Deny-ori.docx yang dibandingkan dengan dengan dokumen uji Deny-x1.docx dengan tingkat kesamaan 80% sampai Deny-x4.docx yang memiliki tingkat kesamaan 20%. Adapun hasilnya ditunjukkan dalam grafik seperti pada Gambar 5.7 sampai Gambar 5.10.



Gambar 5.7. Presentase kesamaan untuk uji coba dokumen Deny-ori.docx.

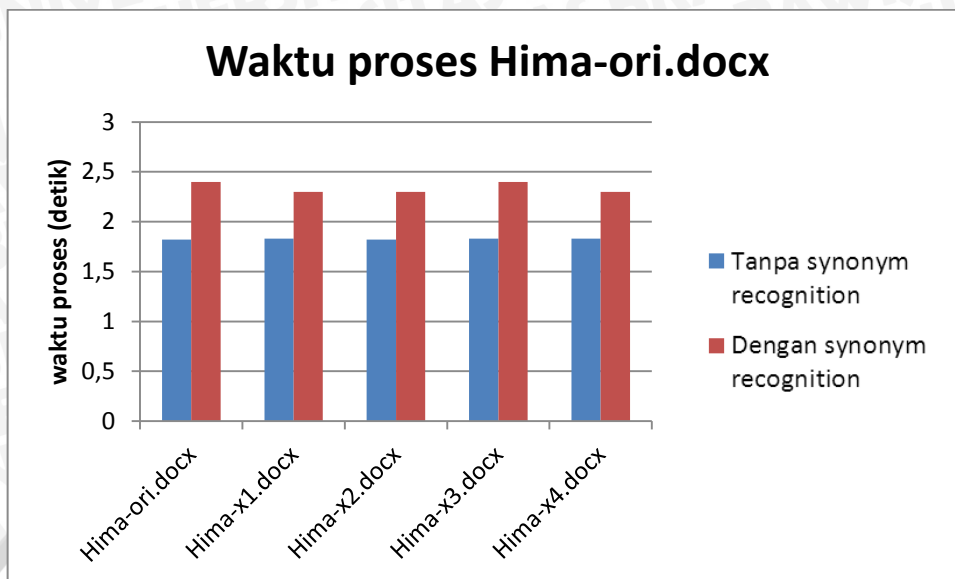


Gambar 5.8. Waktu proses untuk uji coba dokumen Deny-ori.docx.



Gambar 5.9. Presentase kesamaan untuk uji coba dokumen Hima-ori.docx.

Pada Gambar 5.7 dan Gambar 5.9 terlihat bahwa presentase kesamaan *string* antar dokumen berbanding lurus dengan kesamaan kata dalam dokumen uji. Dan penggunaan fitur *synonym recognition* terbukti mempengaruhi presentase kesamaan dengan meningkatkan kecocokan *string*.



Gambar 5.10. Waktu proses untuk uji coba dokumen Hima-ori.docx.

Sedangkan Gambar 5.8 dan 5.10 menunjukkan bahwa waktu proses tidak terpengaruh oleh tingkat kesamaan kata dalam dokumen uji. Waktu proses untuk pencocokan dokumen Deny-ori.docx bahkan relatif sama dengan pencocokan dokumen Hima-ori.docx. Terbukti bahwa banyaknya kata dan struktur kalimat tidak mempengaruhi waktu proses sistem. Meskipun waktu yang dibutuhkan sistem dengan fitur *synonym recognition* lebih tinggi daripada tanpa menggunakan *synonym recognition*. Hal ini dikarenakan dibutuhkan proses tambahan untuk mencari kata ganti di dalam database yang cukup memakan waktu.