

PENGUKURAN KEMIRIPAN DOKUMEN TEKS BAHASA INDONESIA MENGUNAKAN METODE *COSINE* *SIMILARITY*

Dewa Ayu Rai Ariantini, Arie S. M. Lumenta, Agustinus Jacobus.
Teknik Informatika Universitas Sam Ratulangi Manado, Indonesia.
110216022, ayurai15@gmail.com, al@unsrat.ac.id, a.jacobus@unsrat.ac.id

Abstrak - Perkembangan globalisasi pada saat ini dengan teknologi yang sangat pesat dengan perkembangan ilmu pengetahuan dan teknologi kemampuan mahasiswa dibidang ilmu pengetahuan dan teknologi terutama dibidang komputer menjadi salah satu faktor yang memicu tindakan negatif apa bila dilatar belakang berbuat curang, dan rendahnya kemampuan mahasiswa dalam berkreasi dan berinovasi menciptakan karya yang original. Pengolahan informasi dengan menggunakan komputer menghasilkan fasilitas yang *copy paste* sehingga memudahkan pengolahan data sesuai dengan kebutuhan misalnya memenuhi tugas kuliah membuat paper dan sebagainya. Hal ini tentu berpotensi terjadinya tindakan plagiat suatu karya tulis tanpa ijin seperti plagiat. Tahap pengujian sistem dilakukan untuk evaluasi perangkat lunak yang dibangun dengan hasil yang diharapkan, dapat dilakukan dengan cara membandingkan nilai hasil prediksi dan actual, nilai prediksi diperoleh dari sistem, kemudian nilai actual diperoleh dari nilai yang diperiksa secara manual, kemudian nilai yang sudah diketahui prediksi dan actual dengan menggunakan data *dummy* pada sistem deteksi kemiripan dokumen sudah berhasil diterapkan dan sudah bisa digunakan untuk mengukur tingkat kemiripan dokumen.

Kata kunci : *Cosine Similarity*, *Data Dummy*, Plagiat, DFD, *Term Frekuensi*

I. PENDAHULUAN

Era globalisasi ditandai dengan perkembangan ilmu pengetahuan dan Teknologi yang sangat pesat. Kemampuan mahasiswa di bidang ilmu pengetahuan dan teknologi terutama dibidang komputer menjadi salah satu faktor yang memicu tindakan negatif apabila dilatar belakang berbuat curang dan rendahnya kemampuan mahasiswa dalam berkreasi dan berinovasi menciptakan karya yang original. Pengelolaan informasi dengan menggunakan komputer menghasilkan fasilitas yang *copy-paste* (salin-tempel) sehingga memudahkan pengolahan data sesuai dengan kebutuhan misalnya memenuhi tugas kuliah, membuat tugas kuliah dan sebagainya.

Plagiat bukan merupakan fenomena baru dalam dunia pendidikan sebelum hadirnya teknologi informasi, fenomena plagiat juga telah ada namun, hadirnya teknologi informasi secara nyata lebih mempermudah orang untuk melakukan plagiat. Plagiat adalah penjiplakan atau pengambilan karangan, pendapat, dan sebagainya dari orang lain dan menjadikannya seolah karangan dan pendapat sendiri.

Sebenarnya pemerintah sudah memberikan perhatian serius terkait permasalahan plagiat, seperti dalam (Peraturan Menteri Pendidikan Nasional Republik Indonesia nomor 17 tentang pencegahan dan penanggulangan plagiat di perguruan tinggi 2010) dan (Undang-Undang Republik Indonesia nomor 19 tentang hak cipta 2002). Namun pada kenyataannya, tindakan plagiat masih banyak dilakukan di Indonesia dimana tidak hanya melibatkan mahasiswa. Melibatkan dosen yang bergelar Doktor diungkapkan pada Kompas 2013 dan harian Seputar Indonesia, 2012 dalam Suwarjo et al. (2012) yang melaporkan bahwa MZ (Salah satu alumni S3 ITB) telah melakukan plagiat terhadap karya Siyka Zlatanova untuk karya disertasinya.

II. LANDASAN TEORI

A. Pengertian Sistem

Menurut Taufiq (2002), sistem bisa ditafsirkan sebagai kesatuan elemen yang memiliki keterkaitan. Beberapa elemen dapat digabung menjadi suatu unit, kelompok atau komponen sistem dengan fungsi tertentu. Komponen sistem ini bisa dilihat, dianggap atau memang dirancang untuk berfungsi mandiri sebagai modul sistem.

B. *Cosine similarity*

Cosine Similarity adalah ukuran kesamaan antara dua buah vektor dalam sebuah ruang dimensi yang didapat dari nilai cosinus sudut dari perkalian dua buah vektor yang dibandingkan karena cosinus dari 0^0 adalah

1 dan kurang dari 1 untuk nilai sudut yang lain, maka nilai *similarity* dari dua buah vektor dikatakan mirip ketika nilai dari *cosine similarity* adalah 1.

Cosine similarity digunakan dalam ruang positif, dimana hasilnya dibatasi antara nilai 0 dan 1. Kalau nilainya 0 maka dokumen tersebut dikatakan mirip jika hasilnya 1 maka nilai tersebut dikatakan tidak mirip. Perhatikan bahwa batas ini berlaku untuk sejumlah dimensi, dan *Cosine similarity* ini paling sering digunakan dalam ruang positif dimensi tinggi. Misalnya, dalam *Information Retrieval*, masing-masing kata/istilah (*term*) diasumsikan sebagai dimensi yang berbeda dan dokumen ditandai dengan *vector* dimana nilai masing-masing dimensi sesuai dengan berapa istilah muncul dalam dokumen. Berikut adalah rumus *cosine similarity*.

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (1)$$

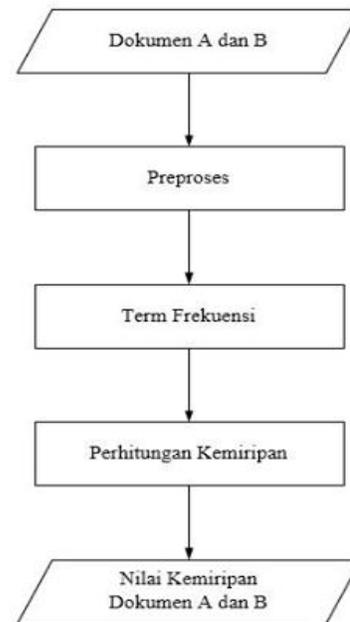
Keterangan :
 A = vektor
 B = vektor
 A_i = bobot *term* i dalam blok A_i
 B_i = bobot *term* i dalam blok B_i
 i = jumlah term dalam kalimat
 n = jumlah vektor

Dimana A merupakan bobot setiap ciri pada *vector* A, dan B merupakan bobot setiap ciri pada B. jika dikaitkan dengan *information retrieval* maka A adalah bobot setiap istilah pada dokumen A, dan B merupakan bobot setiap istilah pada dokumen B. Pada penelitian ini digunakan *cosine similarity* karena citra merupakan salah satu data yang memiliki dimensi tinggi. Pada citra dapat dikatakan bahwa setiap *pixel* merupakan dimensi yang berbeda dan nilai warna pada setiap *pixel* tersebut merupakan nilai dari setiap dimensi. Pengukuran kemiripan dapat dilakukan dengan membandingkan dokumen 1 dengan dokumen 2 kemudian sistem akan menghitung nilai kemiripan. A_iB_i adalah nilai yang diperoleh dari *term* A dan *term* B kemudian kedua nilai tersebut dijumlahkan, kemudian nilai A_i² semua nilai *term* dokumen A semua nilainya dipangkatkan dua, begitu juga dengan *term* B_i² semua nilai yang diperoleh dipangkatkan dua kemudian semua nilai yang diperoleh dijumlahkan.

III. METODOLOGI PENELITIAN

A. Perhitungan Kemiripan

Proses perhitungan kemiripan adalah proses untuk mencari kemiripan antar dokumen. Selanjutnya proses preproses teks adalah untuk menghilangkan tanda baca dan merubah semua huruf menjadi huruf kecil pada dokumen tugas mahasiswa. Kemudian proses sistem menghitung nilai TF atau *term frekuensi*, adalah untuk menghitung kemunculan kata kemudian menyimpannya kedalam basis data tabel indeks. Selanjutnya adalah proses perhitungan kemiripan adalah proses untuk mencari kemiripan antar dokumen teks. kemudian sistem menghitung nilai kemiripan dokumen A dan dokumen B maka user bisa melihat seberapa banyak dokumen yang mirip.



Gambar 1. Perhitungan kemiripan

Proses *upload file* dokumen tugas mahasiswa kemudian disimpan kedalam basis data kemudian sistem preproses teks untuk menghapus tanda baca dalam dokumen, menghilangkan titik, koma, dan sebagainya kemudian sistem menghitung kemiripan TF yaitu *term frekuensi*, adalah proses untuk mengambil setiap kata dari dokumen. Kemudian menghitung nilai kemiripan adalah proses untuk mencari kemiripan antar dokumen. Nilai kemiripan Dokumen A dan dokumen B adalah nilai yang sudah diproses oleh sistem kemudian user bisa melihat nilai kemiripan.

B. Data penelitian

Analisis data dilakukan untuk menganalisa data melakukan perhitungan dengan menggunakan pendekatan *Cosine similarity* sebelum diimplementasikan kedalam sistem, sebelum memasuki tahapan rancangan sistem. Berikut adalah contoh perhitungan kemiripan.

- 1) Perhitungan kemiripan untuk melihat jumlah kata setiap dokumen. Contoh dokumen yang akan di input.
 - Teknik informatika fakultas teknik universitas sam ratulangi manado
 - Program studi teknik informatika jurusan teknik elektro

Disini terdapat dua dokumen yang akan digunakan untuk melakukan pengujian terhadap sistem akan dilihat dari dua dokumen A,B akan melihat kata-kata yang unik pada dokumen A dan B berikut dapat dilihat pada tabel 1.

Nilai dari *term A* dan *term B* dapat diperoleh dari nilai kata yang unik contohnya kata teknik pada *term A* bernilai 2 karena kata teknik *term A* ada 2 begitu juga *term B*, kemudian semua nilai yang diperoleh dari kata yang unik akan dimasukkan dalam rumus *cosine similarity*.

Tabel 1. Perhitungan kemiripan

Tid	Term	T (A)	T (B)
T1	Teknik	2	2
T2	Informatika	1	1
T3	Fakultas	1	0
T4	Universitas	1	0
T5	Sam	1	0
T6	Ratulangi	1	0
T7	Manado	1	0
T8	Program	0	1
T9	Studi	0	1
T10	Jurusan	0	1
T 11	Elektro	0	1

Vektor A dan vektor B tersebut mewakili setiap *term* “A” dan *term* “B” untuk melihat berapa banyak nilai yang diperoleh dalam setiap kata yang unik dalam dokumen, berikut akan dihitung menggunakan metode *cosine similarity*

Vektor A : (2,1,1,1,1,1,1,0,0,0)

B : (2,1,0,0,0,0,0,1,1,1)

Misalkan untuk mencari hasil perhitungan A,B pada tabel 1 perhitungan kemiripan, hasil pencarian adalah sebagai berikut :

$$\text{Similarity (A,B)} = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \cdot \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (2)$$

$$A \cdot B = \sum_{i=1}^n A_i B_i = 4 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 = 5$$

$$|A| = \sum_{i=1}^n A_i^2 = 4 + 1 + 1 + 1 + 1 + 1 + 1 + 0 + 0 + 0 = 10$$

$$|B| = \sum_{i=1}^n B_i^2 = 4 + 1 + 0 + 0 + 0 + 0 + 0 + 1 + 1 + 1 = 9$$

$$= \frac{5}{\sqrt{10} \cdot \sqrt{9}}$$

$$= \frac{5}{9.4}$$

$$= 0,52$$

Ai.Bi adalah nilai yang diperoleh dari *term A* dan *term B*, misalkan *term A* kata “teknik” nilainya 2 dan *term B* kata “teknik” nilainya 2 kemudian dikalikan dan diperoleh nilainya 4,

Ai² adalah nilai yang diperoleh dari *term A* kata “teknik” nilainya 2 kemudian dipangkatkan dan nilainya 4.

Bi² adalah nilai yang diperoleh dari *term B* kata “teknik” nilainya 2 kemudian dipangkatkan dan nilainya 4.

Dengan demikian berdasarkan hasil perhitungan kemiripan dari *term A* dengan *term B* nilai tingkat kemiripan yang diperoleh yaitu 0.52

IV. HASIL DAN PEMBAHASAN

A. Implementasi Proses

Implementasi proses adalah tahapan implementasi dari setiap proses yang sudah dimodelkan pada sistem deteksi kemiripan dokumen. Adapun proses-proses yang sudah dimodelkan adalah proses Tambah data.

- *Proses Tambah Data*

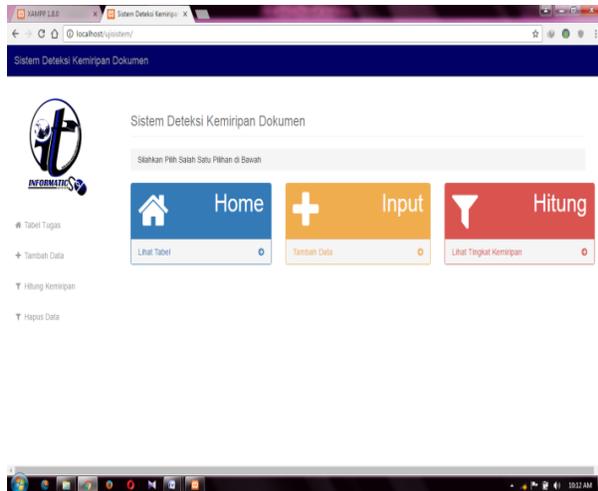
Proses tambah data adalah proses yang ada pada sistem untuk menambahkan data dokumen *file* tugas mahasiswa. Proses tambah data memiliki 4 sub proses yaitu proses *Input* dokumen, *Konfersi PDF* ke teks, *Preprosesing* teks, hitung kemiripan.

- 1) *Implementasi antarmuka*

Implementasi antarmuka menggambarkan tentang hasil implementasi perancangan *storyboard* terhadap perangkat lunak yang dibuat pada sistem deteksi kemiripan dokumen.

- 2) *Antarmuka halaman home*

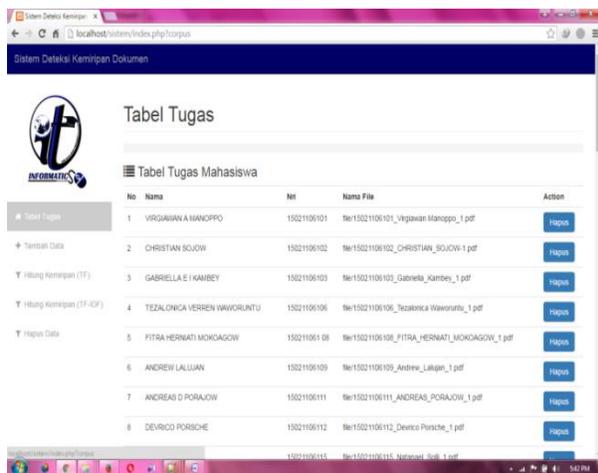
Antarmuka halaman *home* adalah implementasi perancangan antarmuka berdasarkan *storyboard* halaman utama berisi *fitur* yaitu *home*, *input*, dan hitung.



Gambar 2 Antarmuka halaman utama

3) *Antarmuka halaman tabel tugas*

Antarmuka halaman tabel tugas adalah implementasi perancangan antarmuka berdasarkan storyboard halaman tabel tugas mahasiswa dapat dilihat pada gambar 2

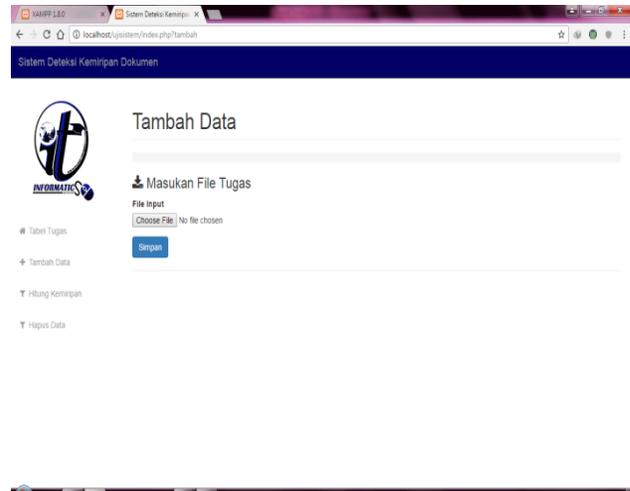


Gambar 3 Antarmuka halaman tabel tugas

Ketika *user* klik tabel tugas mahasiswa akan tampil semua daftar tugas mahasiswa yang sudah di *input*, kemudian disimpan kedalam basis data, kemudian *user* bisa melihat *file* yang sudah diupload.

4) *Antarmuka halaman Tambah data*

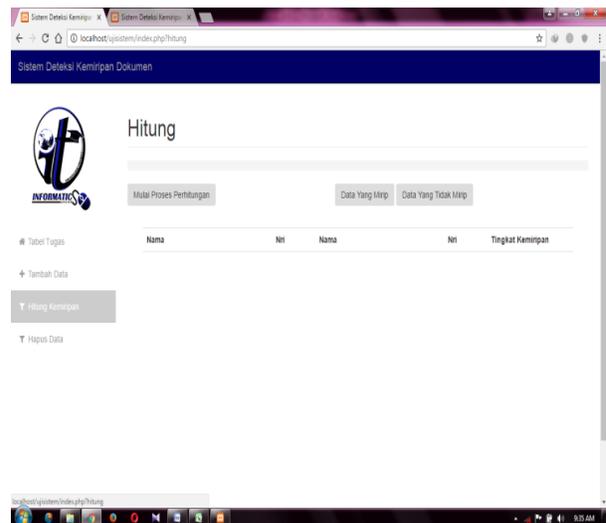
Antarmuka halaman tambah data adalah implementasi dari perancangan antarmuka berdasarkan storyboard halaman tambah data ini berguna untuk upload file tugas mahasiswa.



Gambar 4 Antarmuka halaman Tambah data

5. *Antarmuka halaman Hitung Kemiripan*

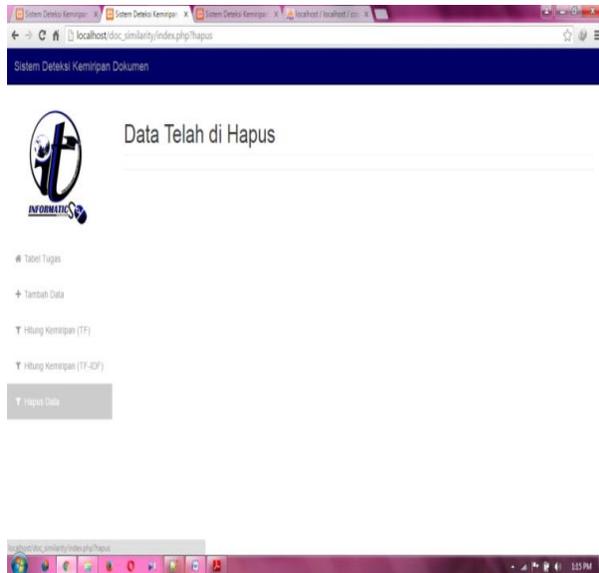
Antarmuka halaman hitung kemiripan adalah implementasi dari Parancangan antarmuka berdasarkan storyboard halaman hitung kemiripan ini berguna untuk menampilkan hasil tingkat kemiripan.



Gambar 5 Antarmuka halaman hitung kemiripan

6. *Halaman Hapus Data*

Antarmuka halaman hapus data adalah implementasi dari Parancangan antarmuka berdasarkan storyboard halaman hapus data ini berguna untuk menghapus dokumen yang sudah diupload.

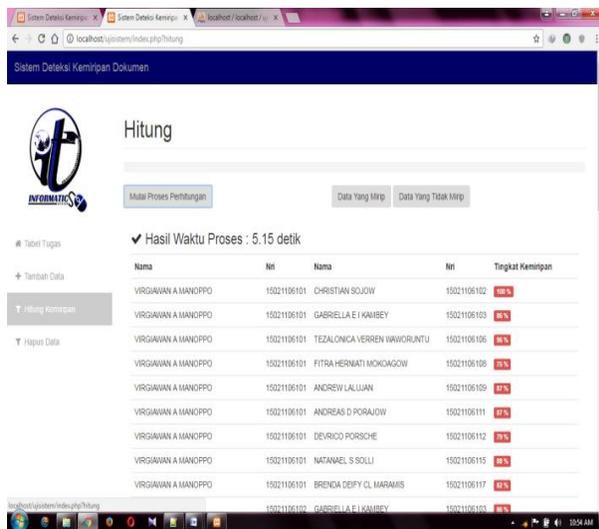


Gambar 6 Antarmuka halaman Hapus data.

B. Pengujian Waktu pengukuran kemiripan

Pengujian ini dilakukan untuk mengetahui waktu pengukuran kemiripan akan diuji berdasarkan banyaknya dokumen yang akan dibandingkan dan lamanya proses Dokumen akan dihitung waktu proses setiap 10 Dokumen untuk lebih jelasnya mengenai hasil pengujian akan dibahas satu per satu di setiap pengujian.

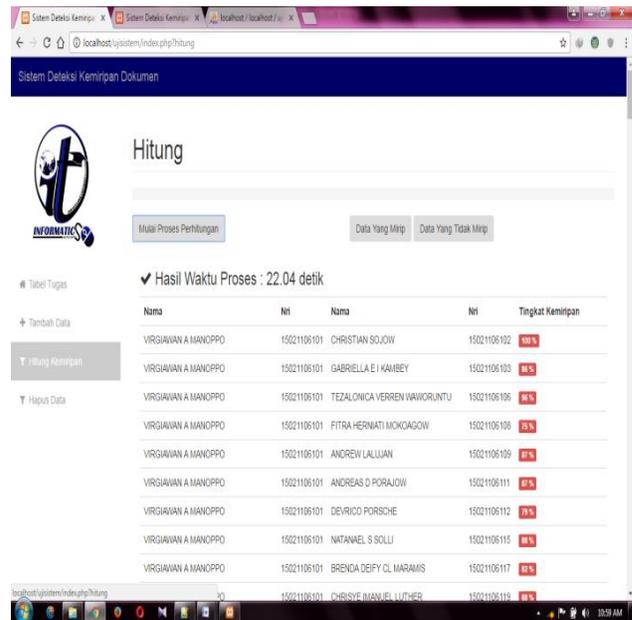
Pengujian dokumen menggunakan 10 data tugas mahasiswa dapat dilihat pada gambar 7



Gambar 7 Pengujian waktu 10 dokumen

Hasil pengujian menggunakan 10 dokumen dapat dilihat pada gambar 7 pengujian waktu menggunakan 10 dokumen. Waktu yang dibutuhkan sistem yaitu 5.15 detik. Dan tidak menutup kemungkinan hasilnya pasti akan berbeda ketika menggunakan komputer yang berbeda.

- Pengujian dokumen menggunakan 20 data tugas mahasiswa dapat dilihat pada gambar 8

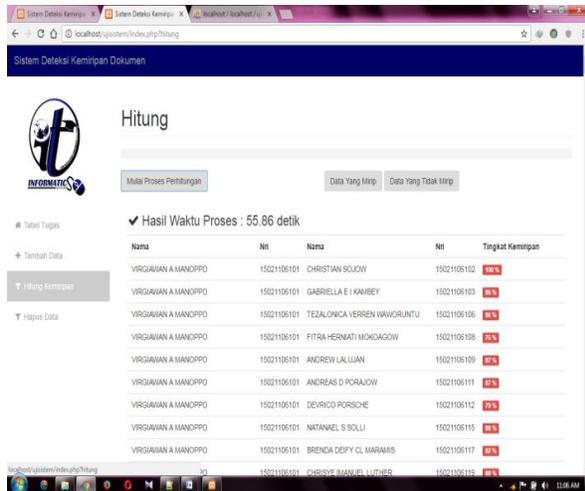


Gambar 8 Pengujian waktu 20 dokumen

Hasil pengujian menggunakan 20 dokumen dapat dilihat pada gambar 8 pengujian waktu menggunakan 20 dokumen. Waktu yang dibutuhkan sistem yaitu 22.04 detik. Pengujian waktu akan terus meningkat seiring dengan banyaknya dokumen yang akan diuji, pengujian waktu dengan 20 data tidak terlalu lama waktu yang dibutuhkan hanya 22.04 detik.

- Pengujian dokumen menggunakan 30 data tugas Mahasiswa

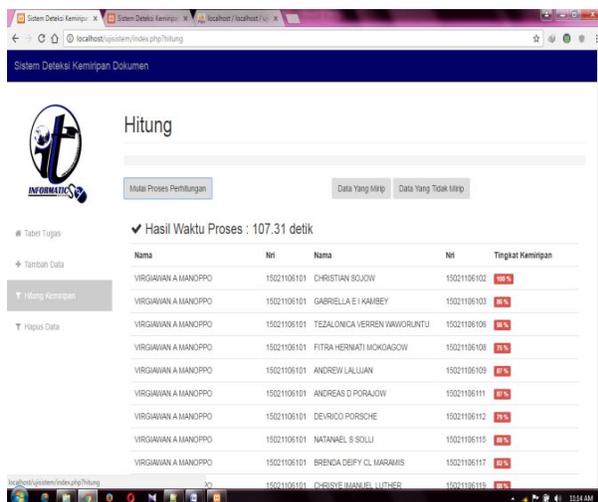
Dapat dilihat pada gambar 9 data yang dibandingkan dalam 30 data tugas mahasiswa sistem memproses data yang sudah diupload kemudian bisa dilihat berapa lama waktu yang dibutuhkan untuk membandingkan 30 dokumen waktu yang dibutuhkan adalah 55.86 detik. Ketika menggunakan spesifikasi berbeda dengan komputer yang lain maka hasilnya juga akan berbeda.



Gambar 9 Pengujian waktu 30 dokumen

Hasil pengujian menggunakan 30 dokumen dapat dilihat pada gambar 9 pengujian menggunakan 30 dokumen. Waktu yang dibutuhkan sistem yaitu 55.86 detik. Pengujian waktu akan terus meningkat seiring dengan banyaknya dokumen yang diuji.

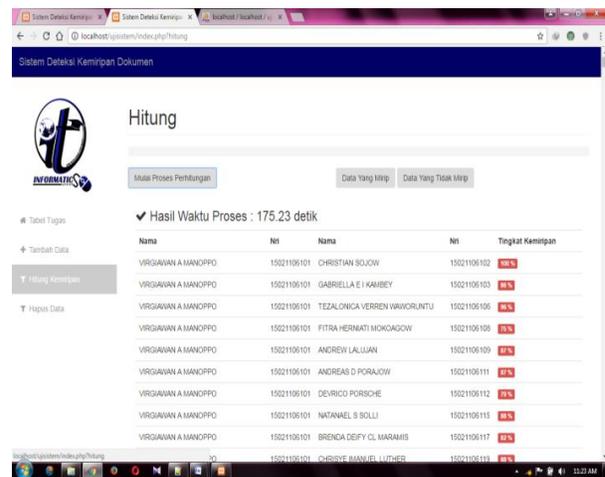
- Pengujian dokumen menggunakan 40 data tugas mahasiswa dapat dilihat pada gambar 10.



Gambar 10 Pengujian waktu 40 dokumen

Hasil pengujian menggunakan 40 dokumen dapat dilihat pada gambar 10 pengujian 40 dokumen. Waktu yang dibutuhkan sistem yaitu 107.31 detik. Pengujian waktu akan terus meningkat seiring dengan banyaknya dokumen yang diuji, pengujian dengan 40 dokumen waktu yang diperlukan cukup lama yaitu 107.31 detik.

- Pengujian dokumen menggunakan 50 data tugas mahasiswa dapat dilihat pada gambar 11 data yang dibandingkan adalah 50 data.



Gambar 11 Pengujian waktu 50 dokumen

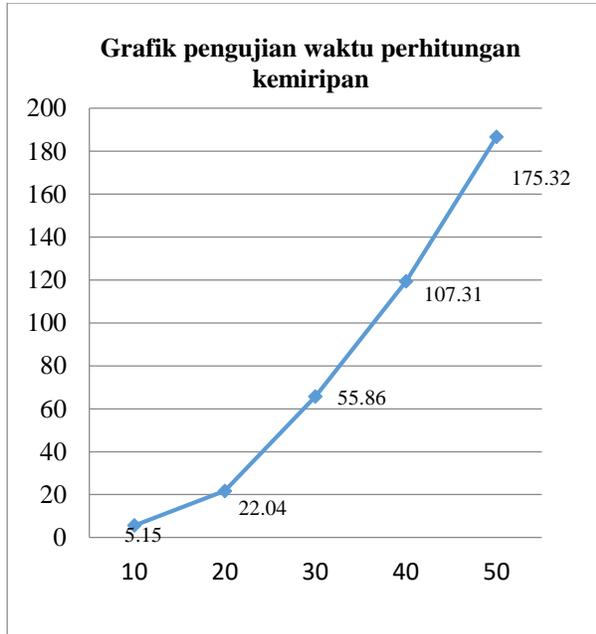
Hasil pengujian menggunakan 50 dokumen dapat dilihat pada gambar 11 pengujian 50 dokumen. Waktu yang dibutuhkan sistem untuk membandingkan 50 dokumen yaitu 175.23 detik. Pengujian waktu akan terus meningkat dengan banyaknya dokumen yang diuji pengujian.

Tabel 2 Pengujian Waktu

No	Bayak Dokumen	Waktu perhitungan kemiripan
1	10	5.15 detik
2	20	22.04 detik
3	30	55.86 detik
4	40	107.31 detik
5	50	175.23 detik

Untuk melihat hasil keseluruhan pengujian waktu perhitungan kemiripan yang telah dilakukan menggunakan data 50 dokumen hasil yang diperoleh dapat dilihat tabel III.

Dapat dilihat pada tabel 2 grafik pengujian waktu perhitungan kemiripan bahwa waktu proses lebih tinggi. Tidak menutup kemungkinan hasil pengujian akan berbeda jika menggunakan spesifikasi komputer yang lain.

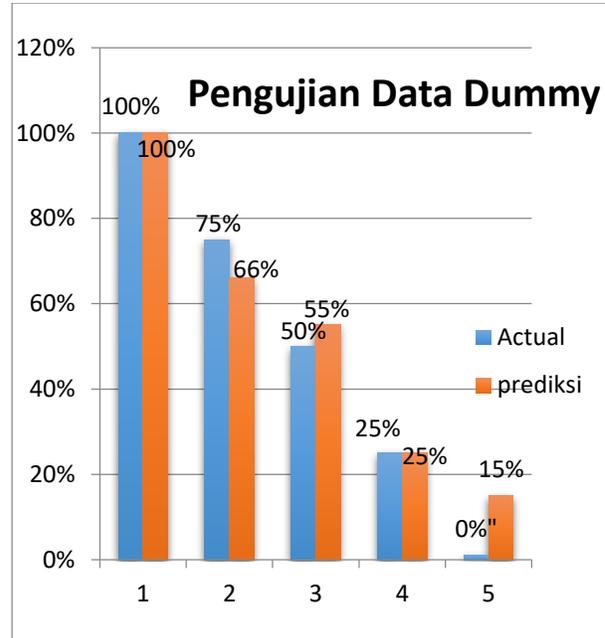


Gambar 12 Pengujian waktu kemiripan

Hasil dari pengujian waktu perhitungan kemiripan bisa dilihat pada gambar 12 pengujian dilakukan menggunakan data tugas mahasiswa waktu yang dibutuhkan tergantung banyaknya dokumen yang akan diuji ketika dokumen terus bertambah maka waktu yang dibutuhkan semakin banyak. *term frekuensi* adalah untuk menyatakan jumlah berapa banyak keberadaan suatu *term* dalam dokumen bisa dilihat pada gambar 12 semakin banyak dokumen yang dibandingkan maka semakin banyak pula waktu yang dibutuhkan.

- Pengujian menggunakan data *dummy*

Data *dummy* adalah data yang digunakan untuk membandingkan dokumen 1 dengan dokumen 2, nilai hasil prediksi dan actual, nilai prediksi diperoleh dari sistem, sedangkan nilai *Actual* diperoleh dari nilai memeriksa tugas mahasiswa secara manual. Berikut adalah pengujian data *dummy* bisa dilihat pada gambar 13.

Gambar 13 Pengujian data *dummy*

Nilai actual dari dokumen satu menjadi acuan nilai actual diperoleh dari memeriksa tugas secara manual kemudian dibandingkan dengan dokumen 2 prediksi dari sistem dengan nilai 100% karena didalam dokumen tersebut memang sama persis. Kemudian dokumen satu dengan dokumen tiga nilai actualnya 66% dan nilai prediksi sebagian isi dokumen diambil dari dokumen satu sehingga nilai yang diperoleh adalah 75%. Nilai actual dokumen satu 55% kemudian nilai prediksi 50% karena sebagian isi dari dokumen tiga diambil dari dokumen dua maka dan juga isi dari dokumen tersebut dikurangi. Nilai actual dokumen satu 25%. Sedangkan nilai prediksi 25% karena didalam dokumen satu sebagian isi diambil dari dokumen 3 dan isi dari tugas dikurangi dan hasil yang diperoleh adalah 25%. Nilai actual pada dokumen satu adalah 1% karena nilai isi dari dokumen tersebut dikurangi. Nilai prediksi dari lima yaitu 15% karena sebagian ini dari dokumen diambil dari dokumen empat dan juga isi dalam dokumen dikurangi. Nilai kesalahan yang diperoleh rata-rata yaitu 7% sehingga masih banyak kesamaan kata atau kesalahan, dengan demikian sistem yang dibangun sudah bisa digunakan untuk mendeteksi kemiripan dokumen.

I. PENUTUP

A. Kesimpulan

Berdasarkan dari pembahasan-pembahasan pada bab-bab sebelumnya, maka dapat ditarik kesimpulan sebagai berikut :

- Sistem untuk mengukur tingkat kemiripan antar dokumen teks berbahasa Indonesia berhasil dibangun menggunakan algoritma *Cosine similarity*.
- Penggunaan data *dummy* pada sistem deteksi kemiripan dokumen sudah berhasil diterapkan dan sudah bisa digunakan untuk mengukur tingkat kemiripan dokumen.

B. Saran

Saran yang ingin disampaikan untuk pengembangan lebih lanjut penelitian ini adalah sebagai berikut :

- Pengembangan bisa membangun sistem *online* agar mempermudah mendeteksi kemiripan dokumen.
- Pengembang nantinya juga bisa mengembangkan sistem agar bisa mendeteksi gambar dari hasil *scan*.

DAFTAR PUSTAKA

- [1] Andriani. *Confusion Matrix Accuracy*. Jakarta. 2013
- [2] Ercegovac, Z. & Jr., J. V. (2012) *Academic Dishonesty, Plagiarism Included, in the Digital Age : A literature review. College & research Libraries, 301-318*
- [3] Hariyanto, Sistem operasi, Bandung, Informatika. 2009.
- [4] J. Hartono, Analisis & Desain , 2005. sistem informasi pendekatan struktur teori Dan praktek aplikasi bisnis. Andi Yogyakarta.
- [5] Undang-undang Republik Indonesia no 19 Tentang Hak Cipta. 2002. Undang-undang Republik Indonesia.
- [6] Pressman, S.R. *SOFTWARE ENGINEERING : A Practitioner's Approach Sixth Edition, New York, Mcgraw-Hill*.2005.
- [7] Kementerian Pendidikan Nasional. 2010, Peraturan Pemerintah Pendidikan Nasional No. 17 tahun 2010 tentang Pencegahan dan Penanggulangan Plagiat di Perguruan Tinggi, diakses pada tanggal 14 November 2014 pada pukul 20.19 WIB dari: <http://luk.tsipil.ugm.ac.id/atur/Permen17-2010.pdf>
- [8] Suwarjo et al. Identifikasi Bentuk Plagiat Pada Skripsi Mahasiswa Fakultas Ilmu Pendidikan Universitas Negeri Yogyakarta”,17. 2012.
- [9] Soeparno. Dasar-dasar Linguistik. Yogyakarta: Mitra Gama Widya. 1993.
- [10] Shutharland Smith. *Plagiarism, the internal and student learning : Improving Academic integrity*. New York : Routledge. 2014.



Sekilas dari penulis dengan nama lengkap Dewa Ayu Rai Ariantini Lahir dikembang mertha, Desa Amertha Buana kec. Dumoga Timur, kabupaten Bolaang Mongondow, Provinsi Sulawesi utara, anak ke-3 dari 4 bersaudara.

Dengan pendidikan Sekolah Dasar Negeri 2 kembang mertha, Kemudian melanjutkan ke Sekolah Menengah Pertama (SMP) Negeri 3 Dumoga, Kemudian melanjutkan ke Sekolah Menengah Kejuruan (SMK) Negeri 3 Manado. Setelah lulus tahun 2011 melanjutkan ke Perguruan Tinggi di Universitas Sam Ratulangi Manado dengan mengambil Jurusan Teknik Informatika. Pada tahun 2016 bulan Oktober, penulis membuat Skripsi demi memenuhi syarat Sarjana (S1) dengan penelitian berjudul Sistem Deteksi Kemiripan Dokumen Teks Bahasa Indonesia Menggunakan Metode *Cosine Similarity*, yang dibimbing oleh dua dosen pembimbing yaitu Arie S. M. Lumenta, ST, MT. dan Agustinus Jacobus, ST, M. Cs. sehingga pada tanggal 3 Oktober 2016 penulis resmi lulus di Teknik Informatika Universitas Sam Ratulangi Manado dan menyandang gelar Sarjana Komputer dengan predikat Sangat Memuaskan.