

**SISTEM TEMU KEMBALI INFORMASI DOKUMEN IDE KREATIF
BERBAHASA INDONESIA MENGGUNAKAN SELEKSI FITUR
DENGAN METODE *INVERSE CLASS SPACE DENSITY*
FREQUENCY (ICS_{δF})**

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:

Riza Aprilia Sari

NIM:125150201111084



**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BRAWIJAYA
MALANG
2016**

PENGESAHAN

SISTEM TEMU KEMBALI INFORMASI DOKUMEN IDE KREATIF BERBAHASA
INDONESIA MENGGUNAKAN SELEKSI FITUR DENGAN METODE *INVERSE CLASS*
SPACE DENSITY FREQUENCY (ICSdF)

SKRIPSI

Diajukan untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun Oleh :

Riza Aprilia Sari

NIM: 125150201111084

Skripsi ini telah diuji dan dinyatakan lulus pada
8 Agustus 2016

Telah diperiksa dan disetujui oleh:

Dosen Pembimbing I

Dosen Pembimbing II

Indriati, S. T, M.Kom

NIP: 19831013 201504 2 002

Rizal Setya Perdana, S.Kom M.Kom

NIK: 201603 910118 1 001

Mengetahui

Ketua Jurusan Teknik Informatika

Tri Astoto Kurniawan, S.T, M.T, Ph.D

NIP: 19710518 200312 1 001

PERNYATAAN ORISINALITAS

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 27 Juli 2016



Riza Aprilia Sari

NIM: 125150201111084

KATA PENGANTAR

Dengan menyebut nama Allah SWT Yang Maha Pengasih dan Penyayang. Puji syukur penulis panjatkan kehadirat Allah SWT karena hanya dengan rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi dengan judul “Sistem Temu Kemali Informasi Dokumen Ide Kreatif Berbahasa Indonesia Menggunakan Seleksi Fitur Dengan Metode Inverse Class Spae Density Frequency”. Shalawat dan salam atas junjungan Nabi besar kita Nabi Muhammad SAW. beserta keluarga dan para sahabat sekalian. Skripsi ini disusun untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana Komputer di Program Studi Informatika/Illmu Komputer Fakultas Ilmu Komputer Universitas Brawijaya Malang.

Melalui kesempatan ini, penulis ingin menyampaikan rasa hormat dan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan bantuan dan dukungan selama penulisan skripsi, diantaranya:

1. Ibu Indriati, S.T, M.Kom selaku dosen pembimbing 1 dan bapak Rizal Setya Perdana, S.Kom, M.Kom selaku dosen pembimbing 2 yang telah membimbing dan mengarahkan penulis sehingga dapat menyelesaikan skripsi ini.
2. Bapak Tri Astoto Kurniawan, S.T, M.T, PhD dan Issa Arwani, S.Kom, M.Sc selaku Ketua dan Sekretaris Program Studi Informatika/Illmu Komputer Universitas Brawijaya.
3. Seluruh dosen Informatika/Illmu Komputer Universitas Brawijaya atas kesediaan membagi ilmunya kepada penulis.
4. Seluruh Civitas Akademika Informatika/Illmu Komputer Universitas Brawijaya yang telah banyak memberi dukungan dan bantuan selama penulis menempuh studi di Informatika/Illmu Komputer Universitas Brawijaya dan selama penyelesaian skripsi ini.
5. Bapak Romli dan Ibu Srinawan selaku orang tua penulis yang telah memberikan doa dukungan moral dan material.
6. Khairana Syasya Mecca ,Momo, Mozi yang telah memberikan motivasi dan semangat demi terselesaikannya skripsi ini.
7. Melly Yansari, Dewi Urifah, Zata Ismah, dan seluruh teman-teman IF-G, serta teman-teman angkatan 2011, angkatan 2012, dan angkatan 2013 yang selalu memberi dukungan, motivasi, kritik, dan saran.

Hanya doa yang penulis bisa berikan semoga Allah SWT memberikan balasan kebaikan yang berlipat. Penulis menyadari bahwa skripsi ini masih banyak kekurangan, oleh karena itu kritik dan saran yang bersifat membangun sangat diharapkan untuk skripsi ini kedepannya. Penulis berharap skripsi ini dapat bermanfaat khususnya bagi diri sendiri dan bagi semua pihak

Malang, 27 Juli 2016

Penulis

rizaaprilia12@gmail.com

UNIVERSITAS BRAWIJAYA



ABSTRAK

Informasi merupakan sekumpulan fakta yang mempunyai arti tertentu. Kebutuhan Informasi yang cepat dan mudah pada saat ini semakin dibutuhkan di segala bidang . Baik pada instansi perusahaan ataupun yang lainnya. Informasi pada dokumen yang dianggap penting dan dibutuhkan sewaktu-waktu tentunya membutuhkan kemudahan dalam pencariannya. Sistem pencarian kata kunci diciptakan pada sebuah dokumen guna mempermudah pencarian Informasi tersebut. Pada kasus ini data dari dokumen tersebut adalah kumpulan ide kreatif pada PJB UP Paiton. Untuk dapat meningkatkan pembangunan PJB UP Paiton melalui saran dan kritik yang terkumpul pada dokumen ide kreatif, dalam skripsi ini dibuatlah program pencarian kata kunci yang diimplementasikan dengan sebuah metode pembobotan yaitu *Inverse Class Space Density Frequency*

Program ini diawali dengan inputan user mengenai kata kunci pada dokumen. Lalu untuk mencari kemiripan *query* yang dimasukkan dengan dokumen dihitung dengan *cosine similarity*. Selanjutnya jarak yang diperoleh dari proses *cosine similarity* digunakan untuk merangking dokumen. Perolehan hasil perankingan selanjutnya digunakan untuk menghitung akurasi dengan memilih dokumen yang relevan. Perolehan akurasi dilakukan dengan menghitung *precision*, *recall* dan *f-measure*. Metode *Inverse Class Space Density Frequency* digunakan untuk pembobotan hasil dari dokumen relevan sesuai kata kunci yang dimasukkan. Perolehan rata-rata akurasi *query* awal 0,833 dan rata-rata akurasi *query* akhir 0,833

Kata kunci: Informasi, ide kreatif, sistem pencarian kata kunci, *Inverse Class Space Density Frequency*, *Cosine Similarity*

ABSTRACT

Information is a collection of facts that have a particular meaning. The information needs quickly and easily at this time are increasingly required in all fields. Neither the agency or other companies. Information documents are considered important and needed at any time certainly requires ease in his quest. Since it was established system of a keyword search on a document in order to facilitate the search of the information. In this case the data of the document is a collection of creative ideas on PJB UP Paiton. In order to enhance the development of PJB UP Paiton through suggestions and criticisms raised in the document creative ideas, in this paper made a keyword search program that is implemented by a weighting method that is Inverse Class Space Density Frequency

The program begins with the user input regarding the keywords in the document. Find similarity query is entered by the document is computed by the cosine similarity. Furthermore, the distance obtained from the similarity cosin used to rank documents. Obtaining the results of the rankings are then used to calculate the accuracy by selecting the relevant documents. Acquisition of accuracy is performed by calculating precision, recall and F-measure. Inverse Method Class Density Frequency Space used for weighting the results of the relevant documents according to the keywords entered by the average. The acquisition of the average accuracy of the initial query is 0,833 and average accuracy of the final query is 0,833

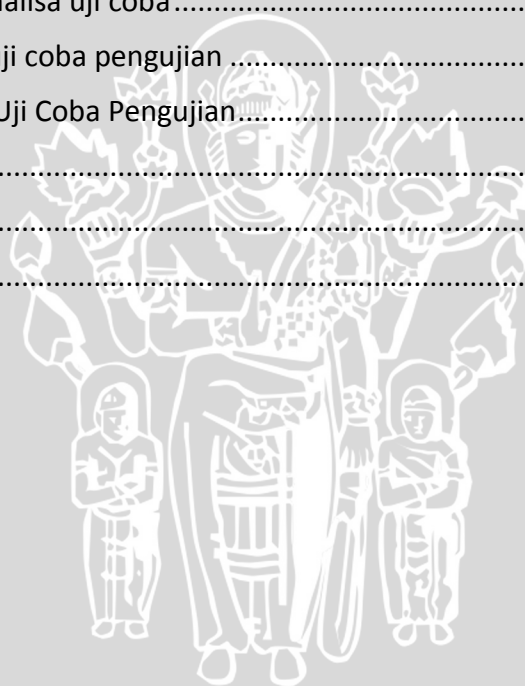
Key words : Information, Creative Ideas, System keyword searches, Inverse Class Space Density Frequency, Cosine Similarity

DAFTAR ISI

PENGESAHAN	ii
PERNYATAAN ORISINALITAS	iii
KATA PENGANTAR.....	iv
ABSTRAK.....	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
DAFTAR SOURCECODE	xv
BAB 1 PENDAHULUAN.....	1
1.1 Latar belakang.....	1
1.2 Rumusan masalah.....	2
1.3 Tujuan	3
1.4 Manfaat.....	3
1.5 Batasan masalah	3
1.6 Sistematika pembahasan	3
BAB 2 LANDASAN TEORI	5
2.1 Tinjauan Penelitian Terdahulu.....	5
2.2 Sistem Temu Kembali Informasi.....	6
2.3 <i>Text Mining</i>	7
2.4 <i>Preprocessing Text</i>	7
2.4.1 <i>Tokenizing</i>	7
2.4.2 <i>Filtering</i>	8
2.4.3 <i>Stemming Porter Algorithm</i>	8
2.5 <i>Vector Space Model</i>	10
2.6 <i>Cosine Similarity</i>	11
2.7 <i>Inverse Class Space Density Frequency (ICS_δF)</i>	12
2.8 <i>Precision, Recall dan F-Measure</i>	14
BAB 3 METODOLOGI PENELITIAN	16
3.1 Tahapan Penelitian	16
3.2 Studi Literatur	16

3.3 Analisis Kebutuhan	17
3.4 Perancangan	17
3.5 Implementasi	18
3.6 Pengujian	18
3.7 Kesimpulan dan Saran	18
BAB 4 Analisis dan PERANCANGAN.....	19
4.1 Teknik Perancangan.....	19
4.1.1 <i>Flowchart</i> Tokenizing	21
4.1.2 <i>Flowchart</i> Filtering	22
4.1.3 <i>Flowchart</i> Stemming	23
4.1.4 <i>Flowchart</i> Cti	25
4.1.5 <i>Flowchart</i> Idf	26
4.1.6 <i>Flowchart</i> Inverse Space Density Space Frequency.....	27
4.1.7 <i>Flowchart</i> cosine similarity.....	29
4.2 Perhitungan Halaman Manualisasi	29
4.3 <i>Preprocessing</i>	30
4.3.1 <i>Tokenizing</i>	30
4.3.2 <i>Filtering</i>	31
4.3.3 <i>Stemming</i>	32
4.3.4 <i>TF.IDF.ICS_δF</i>	32
4.3.5 <i>Cosine Similarity</i>	37
4.4 Desain Antar Muka	48
4.5 Perancangan Pengujian	53
BAB 5 IMPLEMENTASI	54
5.1 Spesifikasi Sistem	54
5.1.1 Spesifikasi Perangkat Keras.....	54
5.1.2 Spesifikasi Perangkat Lunak	54
5.2 Batasan Implementasi Sistem.....	54
5.3 Implementasi Algoritma	55
5.3.1 Implementasi <i>Filtering</i>	55
5.3.2 Implementasi <i>Stemming</i>	56
5.3.3 Implementasi <i>ICS_δf</i>	60

5.3.4 Implementasi <i>Cosine Similarity</i>	63
5.4 Desain Antarmuka	64
5.4.1 Antarmuka Beranda	64
5.4.2 Antarmuka Data	64
5.4.3 Antarmuka <i>Tokenizing</i>	65
5.4.4 Antarmuka <i>Filtering</i>	66
5.4.5 Antarmuka <i>Stemming</i>	67
5.4.6 Antarmuka TF.IDF.ICSDf.....	68
5.4.7 Antarmuka <i>Cosine Similarity</i>	69
5.4.8 Antarmuka Accuracy	70
BAB 6 PENGUJIAN	73
6.1 Hasil dan analisa uji coba	73
6.1.1 Hasil uji coba pengujian	73
6.2 Analisa Uji Coba Pengujian.....	77
BAB 7 KESIMPULAN.....	78
7.1 Kesimpulan.....	78
7.2 Saran	78



DAFTAR TABEL

Tabel 2.1 Aturan <i>Inflectional Particle</i>	8
Tabel 2.2 Aturan <i>Inflectional Possessive Pronoun</i>	9
Tabel 2.3 Aturan <i>First Order Derivational Prefix</i>	9
Tabel 2.4 Aturan <i>Second Order Derivational Prefix</i>	9
Tabel 2.5 Aturan <i>Derivational Suffix</i>	10
Tabel 2.6 <i>Confusion Matrix</i>	15
Tabel 4.1 Dokumen	30
Tabel 4.2 Data hasil <i>tokenizing</i>	30
Tabel 4.3 Data hasil <i>filtering</i>	31
Tabel 4.4 Data hasil <i>stemming</i>	32
Tabel 4.5 Perhitungan <i>TF.IDF.ICSD_F 1</i>	32
Tabel 4.6 Perhitungan <i>TF.IDF.ICSD_F 2</i>	37
Tabel 4.7 Perhitungan <i>TF.IDF.ICSD_F 3</i>	40
Tabel 4.8 Perhitungan <i>Cosine Similarity</i>	43
Tabel 4.10 Hasil <i>cosine</i>	47
Tabel 4.11 Perankingan dokumen	47
Tabel 4.12 Hasil <i>confusion matrix</i>	47
Tabel 4.13 Akurasi	48
Tabel 5.1 Spesifikasi Perangkat Keras	54
Tabel 5.2 Spesifikasi Perangkat Lunak	54
Tabel 5.3 Keterangan Antarmuka Beranda	64
Tabel 5.4 Antarmuka Data	65
Tabel 5.5 Antarmuka <i>Tokenizing</i>	66
Tabel 5.6 Antarmuka <i>Filtering</i>	67
Tabel 5.7 Antarmuka <i>Stemming</i>	68
Tabel 5.8 Antarmuka <i>TF.IDF.ICSD_F</i>	69
Tabel 5.9 Antarmuka <i>Cosine Similarity</i>	70
Tabel 5.10 Antarmuka <i>Accuracy Relevan</i>	71
Tabel 5.11 Antarmuka <i>Accuracy Keseluruhan</i>	71
Tabel 6.1 <i>TF.IDF.ICSD_F precision recall dan f-measure percobaan 1</i>	73
Tabel 6.2 <i>TF.IDF.ICF precision recall dan f-measure percobaan 1.2</i>	74

Tabel 6.3 TF.IDF.ICSDf <i>precision recall</i> dan <i>f-measure</i> percobaan 2	75
Tabel 6.4 TF.IDF.ICF <i>precision recall</i> dan <i>f-measure</i> percobaan 2.1	75
Tabel 6.5 TF.IDF.ICSDf <i>precision recall</i> dan <i>f-measure</i> percobaan 3	76
Tabel 6.6 TF.IDF.ICF <i>precision recall</i> dan <i>f-measure</i> percobaan 3.1	77



DAFTAR GAMBAR

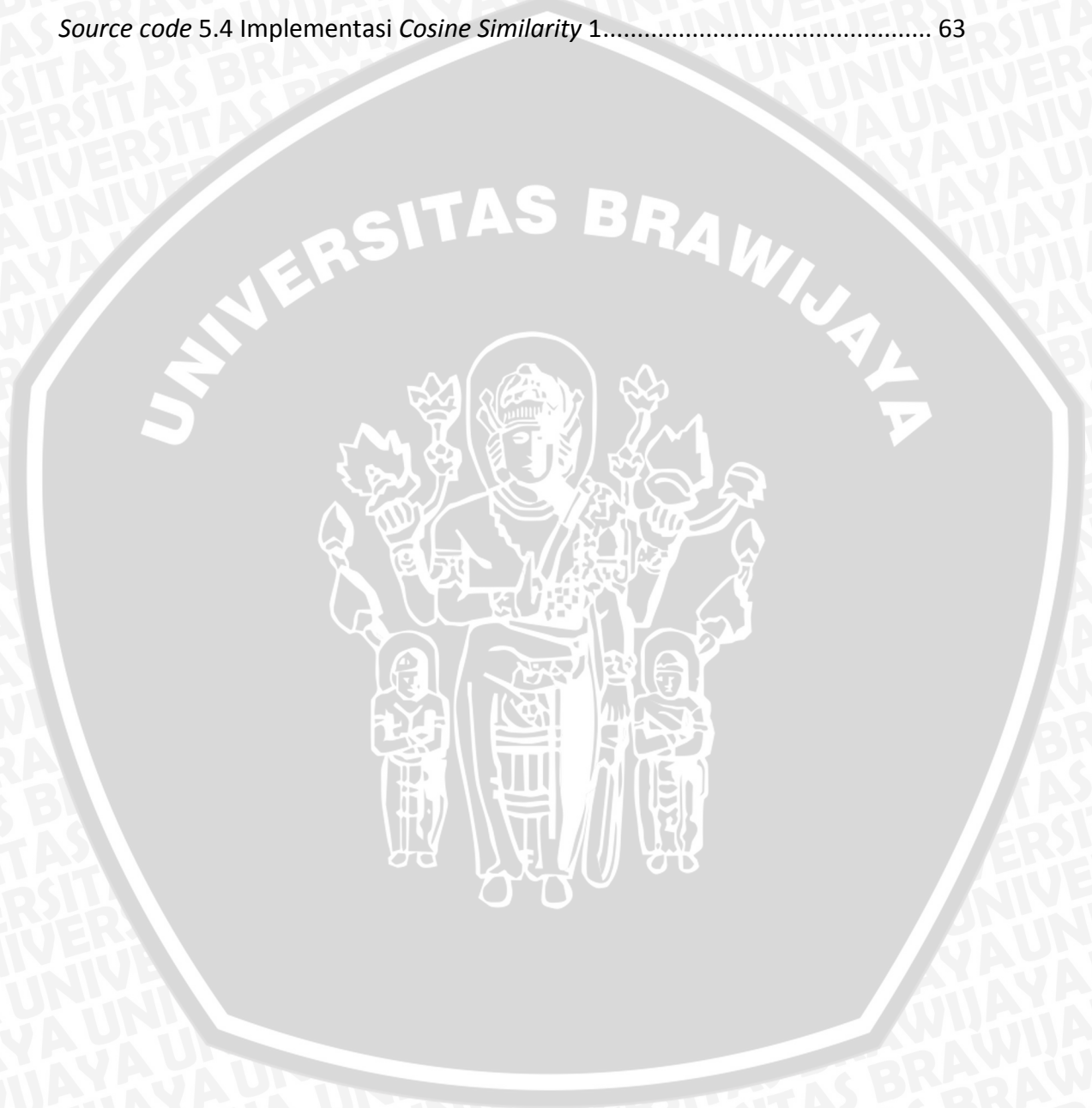
Gambar 2.1 Bagan Sistem Temu Kembali Informasi.....	6
Gambar 2.2 Representasi Perumusan <i>Cosine Similarity</i>	11
Gambar 2.3 <i>Class-oriented-indexing</i> TF.IDF, TF.IDF.ICF dan TF.IDF. ICS ₆ F.....	14
Gambar 3.1 Diagram Blok Metodologi Penelitian	16
Gambar 4.1 <i>Flowchart</i> sistem temu kembali sistem	19
Gambar 4.2 <i>Flowchart Preprocessing</i>	21
Gambar 4.3 <i>Flowchart Tokenizing</i>	22
Gambar 4.4 <i>Flowchart Filtering</i>	23
Gambar 4.5 <i>Flowchart Stemming</i>	25
Gambar 4.6 <i>Flowchart Cti</i>	26
Gambar 4.7 <i>Flowchart Idf</i>	27
Gambar 4.8 <i>Flowchart Inverse Space Density Space Frequency</i>	28
Gambar 4.9 <i>Flowchart cosine similarity</i>	29
Gambar 4.10 Desain Antar Muka Beranda	48
Gambar 4.11 Desain Antar Muka Data	49
Gambar 4.12 Desain Antar <i>Filtering</i>	50
Gambar 4.13 Desain Antar Muka <i>Stemming</i>	50
Gambar 4.14 Desain Antar Muka TF.IDF.ICSdF.....	51
Gambar 4.15 Desain Antar Muka <i>Cosine Similarity</i>	51
Gambar 4.16 Desain Antar Muka Akurasi Relevan	52
Gambar 4.17 Desain Antar Muka Perhitungan Akurasi	52
Gambar 5.1 Antarmuka beranda	64
Gambar 5.2 Antarmuka Data	64
Gambar 5.3 Antarmuka <i>Tokenizing</i>	65
Gambar 5.4 Antarmuka <i>Filtering</i>	66
Gambar 5.5 Antarmuka <i>Stemming</i>	67
Gambar 5.6 Antarmuka TF.IDF.ICSdF	68
Gambar 5.7 Antarmuka <i>Cosine Similarity</i>	69
Gambar 5.8 Antarmuka Accuracy 1	70
Gambar 5.9 Antarmuka <i>Accuracy</i> Keseluruhan	71
Gambar 6.1 Ilustrasi percobaan 1 pengujian 1	74

Gambar 6.2 Ilustrasi percobaan 1.2 pengujian 1	74
Gambar 6.3 Ilustrasi percobaan 2 pengujian 1	75
Gambar 6.4 Ilustrasi percobaan 2.1 pengujian 1	76
Gambar 6.5 Ilustrasi percobaan 3 pengujian 1	76
Gambar 6.6 Ilustrasi percobaan 3.1 pengujian 1	77



DAFTAR SOURCECODE

Source code 5.1 Implementasi <i>Filtering</i>	55
Source code 5.2 Implementasi <i>Stemming</i>	60
Source code 5.3 Implementasi ICSDf	62
Source code 5.4 Implementasi <i>Cosine Similarity 1</i>	63



BAB 1 PENDAHULUAN

Pada bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

1.1 Latar belakang

Dokumen merupakan kumpulan data yang konkrit dan didapatkan berdasarkan proses sistem pengelolaan tertentu. Dokumen juga diartikan sebagai suatu yang tertulis atau tercetak dan segala benda yang memiliki keterangan dipilih untuk dikumpulkan, disusun, disediakan atau disebar. Dengan demikian, dokumen bisa disimpulkan sebagai segala hal baik berupa benda, data, gambar ataupun tulisan yang dipakai sebagai bukti dan tentunya bisa memberikan keterangan yang penting (Poerwadarminta, 2007).

Informasi adalah himpunan fakta atau data yang memiliki makna (Hartono Bambang 2013). Semakin banyaknya kebutuhan akan Informasi dari dokumen yang ada di suatu instansi baik di perguruan tinggi ataupun di perusahaan, tentunya akan semakin menambah kesulitan jika dokumen-dokumen tersebut dicari secara manual. Kecepatan perubahan dan penambahan informasi menyebabkan dibutuhkan suatu sistem yang dapat mengakses dan menyediakan berbagai informasi tersebut. Saat ini telah banyak dari berbagai informasi tersebut dapat diakses secara elektronik melalui *World Wide Web (WWW)* atau internet dengan menggunakan berbagai mesin penelusur. Sistem Temu Kembali Informasi merupakan suatu sistem yang menyimpan informasi dan menemukan kembali informasi tersebut (Saltea, 1983).

Salah satu upaya untuk memaksimalkan sistem temu kembali informasi untuk memudahkan pencarian informasi ini adalah dengan cara pembobotan isi dari dokumen tersebut menggunakan kata kunci. Pembobotan tersebut menggunakan metode *Inverse Class Space density Frequency (ICS_{df})*. Prinsip dari metode ini adalah dengan cara menjumlah hasil dari *Class density term index (C_{dti})* atau kata kunci yang ada pada setiap kelas pada dokumen yang ada, lalu selanjutnya dibagi jumlah kelas nya kemudian di log kan (Fuji Ren dan Mohammad, 2009).

Untuk penelitian sebelumnya yang terkait dengan pembobotan dokumen seperti penelitian yang dibahas pada jurnal yang berjudul "Term Weighting Berbasis Indeks Buku dan Kelas untuk Perangkingan Dokumen Berbahasa Arab". Disini dijelaskan bahwa terdapat beberapa buku berbahasa Arab yang memiliki puluhan bahkan ratusan halaman. *Vector Space Model (VSM)* merupakan tahap yang umum untuk pembobotan dan perangkingan dokumen. Metode yang disebut *Inverse Class Frequency (ICF)* dan *Inverse Book Frequency (IBF)* ini digabungkan dengan metode sebelumnya sehingga menjadi *TF.IDF.ICF.IBF* (Muhammad Ali Fauzi, dkk., 2014). Metode IBF pada jurnal diatas digunakan untuk meningkatkan performa perangkingan dokumen yang memiliki hierarki berupa buku-buku yang memiliki banyak halaman. Sedangkan metode *Inverse Class Space Density Frequency (ICS_{df})* cocok digunakan pada dokumen berbahasa inggris ataupun

berbahasa Indonesia karena terbukti memiliki *precision* dan *recall* yang lebih tinggi daripada TF.IDF (Fuji Ren dan Mohammad, 2013).

Dengan *precision* dan *recall* yang lebih tinggi tentunya tingkat keakurasian Informasi yang diperoleh dari pencarian dokumen akan semakin baik. Dokumen yang akan digunakan pada penelitian ini adalah data Ide Kreatif yang ada pada Perusahaan PJB UP Paiton. Data tersebut berupa kalimat-kalimat Ide atau saran berbahasa Indonesia untuk selanjutnya dipakai sebagai peningkatan pembangunan perusahaan. Sistem pencari ini berguna untuk menemukan data pada dokumen ide kreatif yang tersimpan pada *database*. Mesin pencari menggunakan *query* atau kata kunci pada isi atau fitur digunakan untuk menemukan isi dari dokumen secara spesifik. Sistem ini dibuat diperuntukan kepada admin ataupun anggota lain yang ingin mencari topik untuk ide pada dokumen terkait. Berdasarkan permasalahan-permasalahan yang ada diatas, maka penelitian ini dilakukan untuk membangun sebuah sistem pencari dokumen yang menghasilkan data-data pada dokumen ide kreatif yang mengandung kata kunci. *Inverse class space density frequency* (ICS_sF) dan *Vector Space Model* (VSM). *Vector Space Model* (VSM) adalah salah satu metode pengurutan untuk mencari tingkat kemiripan ide. VSM menggunakan konsep yang terdapat pada aljabar linear yaitu ruang vektor. Berdasarkan konsep yang digunakan tersebut, dikembangkan pemodelan dokumen untuk melakukan pencarian terhadap dokumen yang dibutuhkan. Konsep lain yang digunakan adalah dengan mengetahui kedekatan antara dua buah vektor, dengan cara menghitung besarnya sudut yang terbentuk antara dua vektor dan kemudian diurutkan dari data yang memiliki besar sudut yang terkecil hingga yang terbesar yang menandakan urutan data hasil pengurutan dari yang paling relevan hingga yang tidak relevan (Abdillah Abdul Azis dan Muktyas Indra Bayu, 2012). *Inverse class space density frequency* (ICS_sF) digunakan untuk pembobotan data untuk *query* pada dokumen yang ada. Metode ini cocok digunakan dalam aplikasi berbasis domain seperti klasifikasi teks, pencarian informasi, ekstraksi informasi (Fuji Ren dan Mohammad, 2009).

Dengan adanya sistem untuk pembobotan dari pencarian ide ini diharapkan dapat membantu admin ataupun anggota lain pada perusahaan untuk mencari topik dari banyaknya dokumen yang ada. Alasan inilah yang melatarbelakangi untuk mengambil judul “Sistem Temu Kembali Informasi Dokumen Ide Kreatif Berbahasa Indonesia Menggunakan Seleksi Fitur Dengan Metode *Inverse Class Space Density Frequency* (ICS_sF)”.

1.2 Rumusan masalah

Berdasarkan uraian latar belakang di atas, maka dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana hasil uji coba penerapan metode *Inverse Class Space Density Frequency* dalam Sistem Temu Kembali Informasi untuk dokumen Ide Kreatif?

2. Bagaimana keberagaman tingkat akurasi hasil perhitungan dengan metode *Inverse Class Space Density Frequency* dalam pencarian dokumen Ide Kreatif ?

1.3 Tujuan

Tujuan yang ingin dicapai dalam penelitian ini adalah

1. Menerapkan metode *Inverse Class Space Density Frequency* sebagai uji coba dalam Sistem Temu Kembali Informasi untuk dokumen Ide Kreatif.
2. Menguji tingkat akurasi hasil perhitungan dengan metode *Inverse Class Space Density Frequency* dalam Sistem Temu Kembali Informasi untuk dokumen Ide Kreatif.

1.4 Manfaat

Manfaat yang diperoleh dalam penulisan skripsi ini adalah mencari solusi terbaik guna memudahkan proses pencarian data untuk Dokumen Ide kreatif sesuai dengan *ouput* yang diinginkan.

1.5 Batasan masalah

Untuk menghindari meluasnya masalah yang akan dibahas, maka diberikan batasan masalah sebagai berikut

1. Input berupa *query* atau kata kunci yang relevan sesuai topik yang diinginkan, tahun dipilihnya dokumen yaitu tahun 2014.
2. Output berupa perankingan data dari dokumen yang sesuai dengan kata kunci yang dimasukkan.

1.6 Sistematika pembahasan

Sistematika yang digunakan untuk menyusun laporan penelitian ini adalah sebagai berikut

BAB I PENDAHULUAN

Bab 1 memuat latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan skripsi.

BAB II LANDASAN TEORI

Bab 2 berisi tentang teori-teori yang mendasari dan mendukung penulisan penerapan *Inverse Class Space Density Frequency* Pada Dokumen Ide Kreatif Berbahasa Indonesia.

BAB III METODOLOGI PENELITIAN

Bab 3 berisi langkah-langkah melakukan observasi, pengumpulan dan pengolahan data. Juga akan dibahas *flowchart* mengenai proses-proses dari metode pendukung

BAB IV IMPLEMENTASI

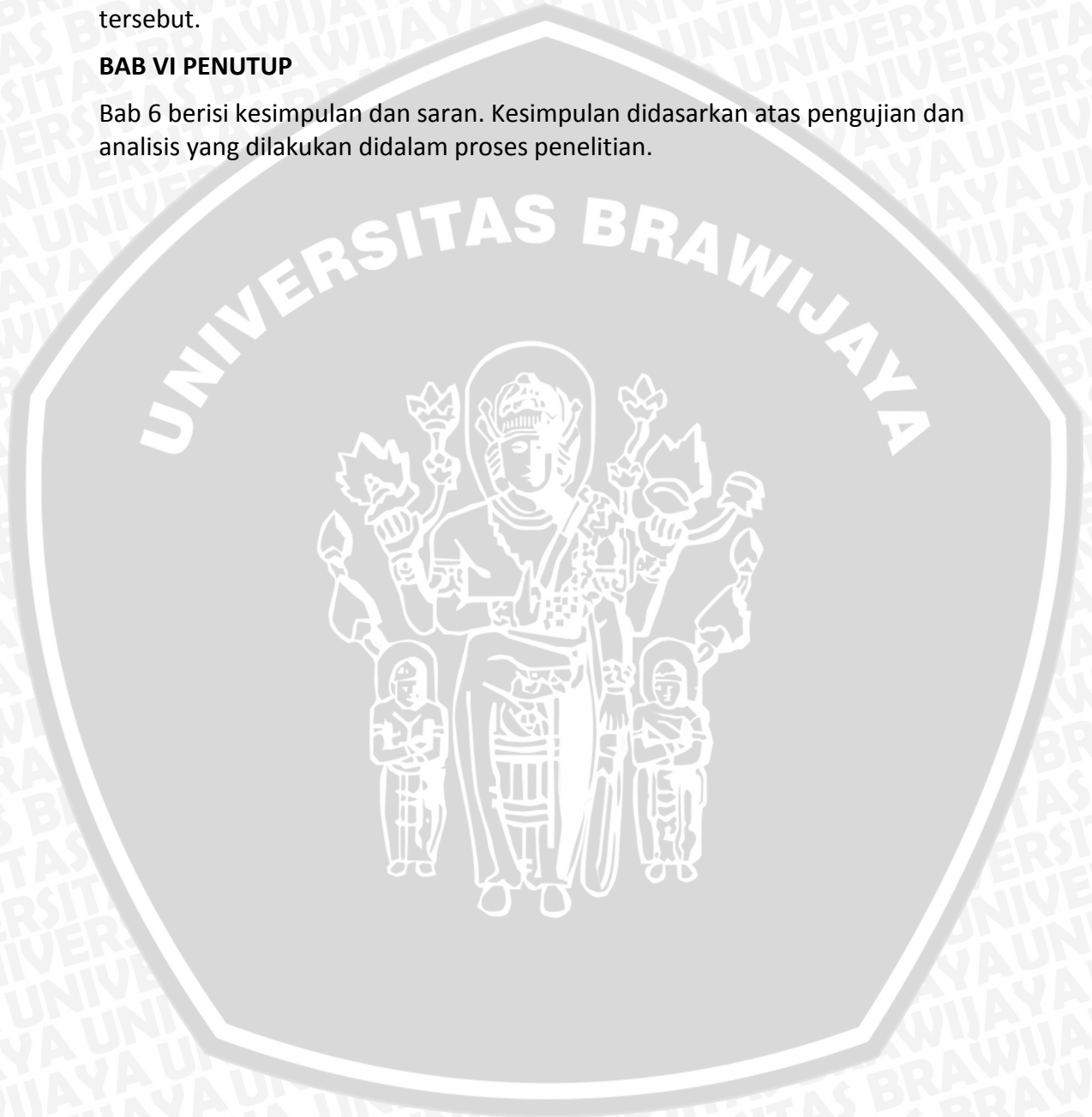
Bab 4 membahas mengenai implementasi dari system dan juga *sourcecode* pada bagian-bagian yang penting.

BAB V PENGUJIAN DAN ANALISIS

Bab 5 berisi tentang proses pengujian dan hasil analisis dari pengujian system tersebut.

BAB VI PENUTUP

Bab 6 berisi kesimpulan dan saran. Kesimpulan didasarkan atas pengujian dan analisis yang dilakukan didalam proses penelitian.



BAB 2 LANDASAN TEORI

Pada bab ini berisi pembahasan review tentang penelitian-penelitian yang terdahulu yang berhubungan dengan permasalahan yang ada dalam penelitian ini. Selain itu pada bab ini juga diuraikan teori-teori pendukung yang digunakan dalam penelitian seperti sistem temu kembali informasi, *text mining*, *vector space model*, *cosine similarity*, *inverse class space density frequency*, *precision*, *recall*, *f-measure*.

2.1 Tinjauan Penelitian Terdahulu

Kajian pustaka dilakukan untuk menganalisa dan membandingkan beberapa penelitian sebelumnya yang berhubungan dengan Sistem Temu Kembali Informasi. Beberapa penelitian yang membahas perangkingan dokumen telah dilakukan sebelumnya, seperti penelitian yang dibahas pada jurnal yang berjudul "Term Weighting Berbasis Indeks Buku dan Kelas untuk Perangkingan Dokumen Berbahasa Arab". Pada jurnal tersebut dijelaskan bahwa terdapat beberapa buku berbahasa Arab yang memiliki puluhan bahkan ratusan halaman. Masing-masing halaman dari buku tersebut adalah sebuah dokumen yang akan diranking berdasarkan query dari pengguna. TF.IDF hanya melakukan pembobotan berbasis pada dokumen tanpa memperhatikan indeks buku dan kelas yang merupakan induk dokumen tersebut sehingga kinerjanya kurang maksimal jika diimplementasikan pada kasus ini. Oleh karena itu, diusulkan metode baru *term weighting* yang berbasis pada indeks buku dan kelas. Metode ini memperhatikan frekuensi kemunculan term pada keseluruhan buku dan kelas. Metode yang disebut *inverse class frequency* (ICF) dan *inverse book frequency* (IBF) ini digabungkan dengan metode sebelumnya sehingga menjadi TF.IDF.ICF.IBF. Untuk data yang dipakai pada penelitian ini merupakan corpus atau kumpulan dokumen teks berbahasa Arab, yang diambil dari 13 kitab dalam perangkat lunak Maktabah Syamilah. Halaman kitab-kitab sebagai suatu dokumen. Jumlah total dokumen dari seluruh kitab tersebut adalah 6996 dokumen yang tersebar dalam 5 kategori. Dan dari seluruh dokumen dataset tersebut terdapat 47.447 kata berbeda (*distinct term*). Berdasarkan uji coba yang telah dilakukan didapatkan nilai *F-measure* sebesar 75%, rata-rata *precision* 76% dan *recall* mencapai 74%. TF.IDF.ICF.IBF cocok diaplikasikan pada perangkingan dokumen pada buku-buku yang memiliki banyak halaman (Muhammad Ali Fauzi, dkk., 2014).

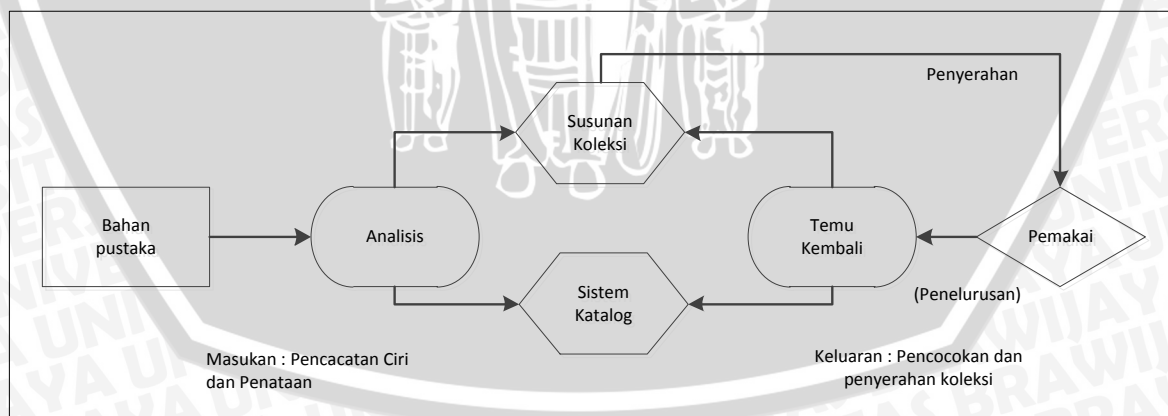
Untuk penelitian yang kedua berjudul *Query Expansion* Pada Sistem Temu Kembali Informasi Dokumen Berbahasa Indonesia Menggunakan *Pseudo Relevan Feedback*. Penelitian ini membahas tentang bagaimana *user* kesulitan dalam menemukan suatu dokumen dengan *query* yang kurang spesifik.

Berdasarkan permasalahan tersebut dibutuhkan suatu metode untuk memperbaiki *query* yang dimasukkan oleh *user* sehingga dapat meningkatkan kinerja dari sistem pencarian dan meningkatkan kepuasan *user*. *Query expansion* yaitu menambahkan *query* tambahan pada *query* awal agar performa ketepatan meningkat dan diharapkan dapat memberikan informasi yang lebih akurat kepada

user. Perluasan *query* dapat dilakukan dengan metode *pseudo relevance feedback* yang merupakan teknik ekspansi *query* yang mengekstraksi istilah yang sangat berbobot dari dokumen peringkat teratas yang dikembalikan oleh pencarian awal. Data pada penelitian ini menggunakan 37.765 koleksi *local content* yang terdiri dari skripsi, disertasi dan tesis pada Perpustakaan Universitas Brawijaya. Pengujian pada penelitian ini dilakukan dengan dua skenario dan beda *query* untuk masing-masing percobaannya. Skenario pertama meliputi percobaan 1-5 Skenario kedua meliputi percobaan 6-7. Untuk keseluruhan hasil pengujian *Query Expansion* Pada Sistem Temu Kembali Informasi Dokumen Berbahasa Indonesia Menggunakan *Pseudo Relevan Feedback* ini menghasilkan rata-rata *precision* 0,17, kenaikan rata-rata *recall* sebesar 0,03 dan kenaikan rata-rata *f-measure* sebesar 0,18 (Zanwar, 2015).

2.2 Sistem Temu Kembali Informasi

Temu kembali informasi atau *information retrieval* merupakan proses dimana pengguna dapat menemukan informasi yang dibutuhkan pada penyedia informasi dengan dibantu oleh sistem yang sudah disediakan. Menurut Saltea 1983, dalam Janu Saptari dan Purwono 2006 menjelaskan bahwa secara sederhana temu kembali informasi merupakan suatu sistem yang menyimpan informasi dan menemukan kembali informasi tersebut (Janu Saptari dan Purwono, 2006). Pada dasarnya sistem temu kembali informasi yang bertujuan untuk menyimpan informasi adalah sebuah kumpulan laporan yang tersimpan secara bersama-sama dalam satu tempat penyimpanan. Laporan-laporan yang tersimpan dapat berbentuk bibliografi koleksi yang berada di penyedia jasa tersebut, bibliografi dari koleksi tersebut digunakan sebagai bahasa penelusur informasi. Pada bagan yang dibuat oleh Lauren B. Doyle dalam Miswan (2003), Juga terdapat kemiripan pada bagan sistem temu kembali informasi milik Lancaster, berikut ini Gambar 2.1 Lauren B. Doyle:



Gambar 2.1 Bagan Sistem Temu Kembali Informasi

Sumber: Lauren B. Doyle

Masukan merupakan kegiatan yang dilakukan oleh perpustakaan, yaitu semua, bahan pustaka atau rekaman informasi diorganisasir, diolah, di katalog di

klasifikasi (analisis) yang menghasilkan susunan bahan pustaka di rak (susunan koleksi) dan wakil ringkas bahan pustaka yang berupa katalog, bibliografi, indeks. Sedangkan keluaran adalah kegiatan temu kembali informasi oleh pemakai perpustakaan. Dalam temu kembali informasi di perpustakaan, pemakai dapat menempuh dua cara, yaitu langsung menuju ke susunan koleksi di rak atau melalui sistem katalog baru menuju ke rak. Menurut Wibowo (2012) Sistem Temu Balik Informasi (Information Retrieval) adalah ilmu mencari informasi dalam suatu dokumen, mencari dokumen itu sendiri dan mencari metadata yang menggambarkan suatu dokumen. Sistem Temu Balik Informasi merupakan cabang dari ilmu komputer terapan (*applied computer science*) yang berkonsentrasi pada representasi, penyimpanan, pengorganisasian, akses dan distribusi informasi. Dalam sudut pandang pengguna, Sistem Temu Balik Informasi membantu pencarian informasi dengan memberikan koleksi informasi yang sesuai dengan kebutuhan pengguna.

2.3 Text Mining

Text mining termasuk wilayah penelitian baru di ilmu komputer yang menarik yang mencoba menemukan solusi akibat *overload* informasi dengan menggabungkan teknik dari *data mining*, pembelajaran mesin (*machine learning*), pengolahan bahasa alami (*natural language processing*), pengambilan informasi (*information retrieval*), dan manajemen pengetahuan (*knowledge management*) (Feldman, dkk., 2007).

Manfaat dari *text mining* adalah untuk mempermudah pencarian dan membuat inovasi yang dapat membantu manusia untuk mengerti dan menggunakan informasi dari sebuah repositori dokumen (Srivastava, dkk., 2009). Di dalam proses *text mining* dilakukan beberapa tahapan umum diantaranya adalah *tokenizing*, *filtering*, *stemming* dan *analyzing*. Ketiga tahap awal (*tokenizing*, *filtering* dan *stemming*) merupakan tahap *preprocessing text*. Tahap terakhir dalam *text mining* adalah tahap *analyzing* yaitu tahap penentuan seberapa jauh keterhubungan antar kata-kata antar dokumen yang sebelumnya dilakukan pembobotan kata dan representasi dokumen ke vektor.

2.4 Preprocessing Text

Preprocessing Text merupakan tahap awal *text mining* yang terdiri diantaranya adalah *tokenizing*, *filtering*, dan *stemming*. Tahap ini bertujuan untuk mempersiapkan data dokumen agar bisa menjadi data terstruktur dan berupa nilai numerik maka data dapat dijadikan sebagai sumber data yang dapat diolah lebih lanjut.

2.4.1 Tokenizing

Tahap *tokenizing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya (Garcia, 2005) Kata dan istilah sederhana itu berupa potongan-potongan kata tunggal yang menyusun suatu dokumen. Pada tahap ini, dilakukan

pemotongan (*parsing*) terhadap kata-kata tunggal tersebut menjadi kumpulan *token*.

2.4.2 Filtering

Filtering adalah proses menentukan kata-kata (*term-term*) apa saja yang akan digunakan untuk mempresentasikan dokumen. Selain untuk menggambarkan isi dokumen, *term* ini juga berguna untuk membedakan dokumen yang satu dengan dokumen lainnya pada koleksi dokumen (Garcia, 2005). Proses ini dilakukan dengan mengambil kata-kata penting dari hasil *token* dan menghapus *stop words*. *Stop words* adalah kata-kata yang bukan merupakan ciri (kata unik) dari suatu dokumen seperti kata sambung, kata kepemilikan. Memperhitungkan *stopword* pada transformasi teks akan membuat keseluruhan sistem *text mining* bergantung kepada faktor bahasa. Hal ini menjadi kelemahan dari proses penghilangan *stopword* (Tala dan Fadillah, 2003). Namun proses penghilangan *stop word* tetap digunakan karena proses ini akan sangat mengurangi beban kerja sistem. Dengan menghilangkan *stopword* dari suatu teks maka sistem hanya akan memperhitungkan kata-kata yang dianggap penting.

2.4.3 Stemming Porter Algorithm

Stemming adalah suatu proses yang menyediakan suatu pemetaan antara berbagai kata dengan morfologi yang berbeda menjadi satu bentuk dasar (*stem*) (Tala dan Fadillah, 2003). Penelitian ini menggunakan Algoritma Porter. Algoritma ini mempunyai langkah-langkah seperti berikut ini (Ledy Agusta, 2009):

1. Hapus *Particle*
2. Hapus *Possesive Pronoun*
3. Hapus awalan pertama. Jika tidak ada lanjutkan ke langkah 4a, jika ada cari maka lanjutkan ke langkah 4b
4. a. Hapus awalan kedua, lanjutkan ke langkah 5a.
b. Hapus akhiran, jika tidak ditemukan maka kata tersebut diasumsikan sebagai *root word*. Jika ditemukan maka lanjutkan ke langkah 5b.
5. a. Hapus akhiran. Kemudian kata akhir diasumsikan sebagai *root word*
b. Hapus awalan kedua. Kemudian kata akhir diasumsikan sebagai *root word*

Terdapat 5 kelompok aturan pada Algoritma Porter untuk Bahasa Indonesia ini. Aturan tersebut dapat dilihat pada Tabel 2.1 sampai Tabel 2.5 dibawah ini :

Tabel 2.1 Aturan Inflectional Particle

Akhiran	Replacement	Additional Conditional
-kah	Null	Null
-lah	Null	Null
-pun	Null	Null

Tabel 2.2 Aturan *Inflectional Possessive Pronoun*

Akhiran	Replacement	Additional Conditional
-ku	Null	Null
-mu	Null	Null
-nya	Null	Null

Tabel 2.3 Aturan *First Order Derivational Prefix*

Awalan	Replacement	Additional Conditional
Meng-	Null	Null
Meny-	S	V...*
Men	Null	Null
Mem	P	V...
Me-	Null	Null
Peng-	Null	Null
Peny-	S	V...
Pen-	Null	Null
Pem-	P	V...
Pem-	Null	Null
di-	Null	Null
Ter-	Null	Null
Ke-	Null	Null
Mem-	Null	Null

Tabel 2.4 Aturan *Second Order Derivational Prefix*

Awalan	Replacement	Additional Condition
Ber-	Null	Null
Bel-	Null	Ajar
Be	Null	K*er
Per-	Null	Null
Pel-	Null	Ajar
Pe-	Null	Null

Tabel 2.5 Aturan *Derivational Suffix*

Akhiran	Replacement	Additional Condition
-kan	Null	Prefix bukan anggota {ke,pang}
-an	Null	Prefix bukan anggota {di, meng, ter}
-i	Null	Prefix bukan anggota {ber, ke, peng}

Sumber : Ledy Augusta.

2.5 Vector Space Model

Vector Space Model (VSM) mempresentasikan setiap dokumen yang terdapat dalam *database* dan *query* ke dalam vektor multidimensi. Dimensi dari vektor sesuai dengan jumlah setiap *term* dalam *database* dan kumpulan *term* tersebut membentuk suatu ruang vektor (Tala dan Fadillah, 2003).

Dokumen dan *queries* mewakili vektor masing-masing, seperti pada VSM setiap *term*, *i*, di dalam dokumen maupun *query*, *j*, diberikan suatu bobot yang bernilai *real* w_{ij} . Dokumen dan query diekspresikan sebagai vektor *t*- dimensi

$$d_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j}) \quad (2.1)$$

Keterangan

d_j : dokumen ke- *j*

$w_{1,j}, w_{2,j}$: bobot query ke 1,2

jika terdapat *n* dokumen di dalam database, yaitu $j = 1, 2, \dots, n$. Contoh dari VSM dengan tiga dimensi untuk dua dokumen D_1 dan D_2 , satu *query* *Q*, dan tiga *term* T_1, T_2 , dan T_3 .

Selain itu pada VSM, *database* dari semua dokumen wakili oleh *matriks term document* atau (*matriks term-frequency*). Dimana setiap sel pada *matriks* sesuai dengan bobot yang diberikan dari suatu *term* dalam dokumen yang ditentukan. Nilai nol berarti bahwa *term* tidak terdapat dalam dokumen tersebut.

Inverse Document Frequency (IDF). IDF didefinisikan sebagai persamaan berikut

$$idf_i = \log \left(\frac{n}{df_i} \right) \quad (2.2)$$

Keterangan

idf_i = *Inverse document frequency* untuk tiap *term* atau kata.

log = Operasi matematika yang merupakan kebalikan dari operasi pangkat.

n = Jumlah dokumen yang ada pada data training.

df_i = Jumlah dokumen yang memiliki kata tersebut.

df_i merupakan frekuensi dokumen dari *term i* atau sama dengan jumlah dokumen yang mengandung *term i* dan n adalah total dokumen di dalam database. *Log* digunakan untuk memperkecil pengaruh *relative* untuk tf_{ij} .

Bobot w_{ij} dihitung menggunakan ukuran TF.IDF (*term frequency-inversed document frequency*) didefinisikan sebagai persamaan berikut ini :

$$w_{ij} = tf_{ij} \times idf_i \quad (2.3)$$

Keterangan :

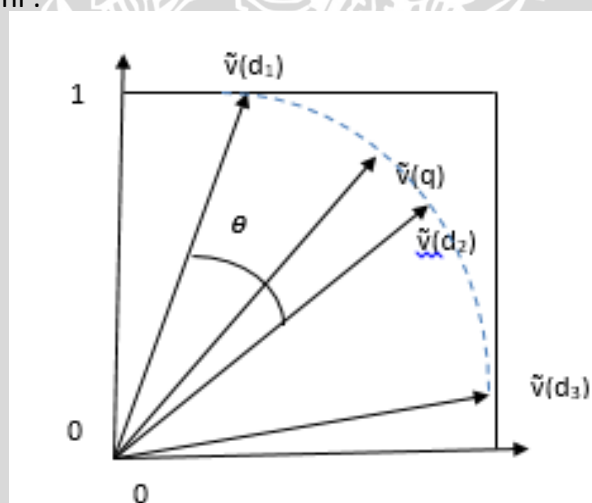
w_{ij} = Hasil perkalian dari tf_{ij} dan idf_i .

tf_{ij} = Frekuensi kemunculan kata atau *term* pada dokumen.

idf_i = *Inverse document frequency* untuk tiap *term* atau kata.

2.6 Cosine Similarity

Hasil pembobotan kata pada dokumen digunakan sebagai representasi vektor. Dari representasi bobot tersebut dapat dihitung nilai kemiripan suatu dokumen dengan *query*. Nilai kemiripan ini biasa dihitung dengan rumusan *cosine similarity*, perhitungan tingkat kemiripan ini dibuat dengan berdasar pada besar sudut kosinus antara dua vektor, dalam hal ini adalah vektor dokumen. Representasi perumusan ini dalam bidang kartesian seperti diperlihatkan pada Gambar 2.2 berikut ini :



Gambar 2.2 Representasi Perumusan *Cosine Similarity*

Sumber : Krzysztof J. Cios,

Dalam Gambar 2 terdapat tiga vektor dokumen d_1 , d_2 dan d_3 dan satu vektor *query q*. *Cosine similarity* menghitung nilai kosinus θ dari *query* dan tiga dokumen lain. Nilai ini menunjukkan derajat kemiripan dokumen dengan *query*. Karena berdasarkan kosinus sudut antara dua vektor, maka nilainya berkisar pada 0 sampai dengan 1, dimana 0 menandakan bahwa kedua dokumen tidak mirip sama sekali, dan 1 menandakan bahwa antara *query* dan dokumen benar-benar identik. *Cosine* dinyatakan sebagai berikut

$$\cos(q, d_j) = \frac{\sum_{tk} [TFIDF(t_k, q)] \cdot [TFIDF(t_k, d_j)]}{\sqrt{\sum_i |TFIDFq_i|^2} \cdot \sqrt{\sum_j |TFIDFd_j|^2}} \quad (2.4)$$

Keterangan :

$\cos(q, d_j)$: merupakan nilai kosinus antara *query* dan dokumen *j*

$TFIDF(t_k, q)$: pembobotan TFIDF kata *term* ke *k* pada *query*

$TFIDF(t_k, d_j)$: pembobotan TFIDF kata *term* ke *k* pada *query* dan dokumen *j*

$|TFIDFq_i|$: panjang dari vektor *query* ke *i*

$|TFIDFd_j|$: panjang dari vektor dokumen ke *j*

Salah satu ukuran kemiripan teks yang populer digunakan pada VSM untuk pencarian dokumen adalah cosine similarity (Krzysztof J. Cios, 2007).

Konsep dari *cosine similarity* yaitu menghitung nilai *cosinus* sudut antara dua vektor yaitu jika diberikan dokumen yang diwakili oleh vektor d_j dan *query* q , dan *term* yang diekstrak dari *database*, maka nilai *cosine similarity* didefinisikan sebagai berikut

$$Sim(q, d_j) = \left(\frac{d_j \cdot q}{|d_j| \cdot |q|} \right) = \frac{\sum_{i=1}^t W_{iq} \cdot W_{ij}}{\sqrt{\sum_{j=1}^t (W_{iq})^2 \cdot \sum_{i=1}^t (W_{ij})^2}} \quad (2.5)$$

Keterangan

$|q|$: panjang *query*

$|d_j|$: panjang dokumen

W_{iq} : bobot *term* *i* pada *query*

W_{ij} : bobot *term* *i* pada dokumen

Sudut yang merentang antara vektor d_j dan *query* q akan menghasilkan sudut yang jika semakin kecil sudut diantara kedua vektor d_j dan *query* q , maka akan semakin tinggi derajat kesamaan. *Cosinus* dari sudut tersebut merupakan koefisien yang dapat mewakili kemiripan antara vektor d_j dan *query* q (Jovita, dkk., 2015)

2.7 Inverse Class Space Density Frequency (ICS_{DF})

Class space density untuk *term* t_i dan kelas lain dirumuskan sebagai berikut

$$CS\delta(t_i) = \sum_{c_k} C(t_i) \quad (2.6)$$

Keterangan :

$CS\delta(t_i)$: *Class Space Density* *term* ke *i*

c_k : Certain Category

C : Jenis kelas pada dokumen

Untuk *inverse class space density frequency* sebagai berikut :

$$ICS\delta F(t_i) = \log \left(\frac{c}{CS\delta} \right) \quad (2.7)$$

Keterangan

$ICS\delta F(t_i)$: *inverse class space density frequency* term ke i

C : jenis kelas pada dokumen

$CS\delta$: *Class Space Density*

Representasi numerik dari frekuensi *term* (*local* parameter), IDF (*global* parameter) dan di *inverse class space density frequency* dinyatakan sebagai rumus $TF.IDF.IC\delta F$ untuk term t_i pada dokumen d_j dengan kategori c_k didefinisikan sebagai berikut

$$W_{TF.IDF.IC\delta F}(t_i, d_j, C_k) = tf(t_i, d_j) \times \left(1 + \log \frac{D}{d(t_i)} \right) \times \left(1 + \log \frac{c}{CS\delta(t_i)} \right) \quad (2.8)$$

Keterangan

$W_{TF.IDF.IC\delta F}$: bobot *term frequency*, *Index document frequency*, *inverse class space density frequency*

t_i : *term* ke i

d_j : dokumen ke j

c_k : *Certain Category*

D : jumlah keseluruhan dokumen

$d(t_i)$: dokumen *term* ke i

C : jenis kelas pada dokumen

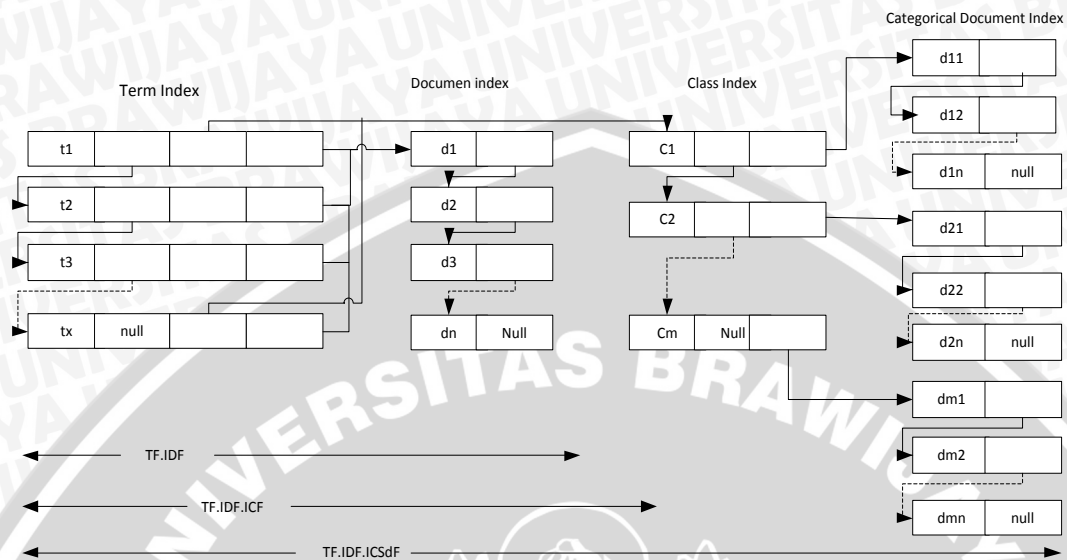
$CS\delta(t_i)$: *Class Space Density term i*

Pada normalisasi *cosine* didapatkan rumus

$$W_{TF.IDF.IC\delta F}(t_i, d_j, C_k) = \frac{tf(t_i, d_j) \times \left(1 + \log \frac{D}{d(t_i)} \right) \times \left(1 + \log \frac{c}{CS\delta(t_i)} \right)}{\sqrt{\sum_{t_i \in d_j, t_i \in c_k} [tf(t_i, d_j) \times \left(1 + \log \frac{D}{d(t_i)} \right) \times \left(1 + \log \frac{c}{CS\delta(t_i)} \right)]^2}} \quad (2.9)$$

Dimana $\frac{CS\delta(t_i)}{c}$ mengacu pada *class space density frequency* ($CS\delta F$) dan $\frac{c}{CS\delta}$ adalah *inverse space density frequency* ($ICS\delta$) dari *term* t_i . Penggunaan pembobotan berbasis kelas untuk *term weighting* pada dokumen berbahasa Inggris yang dinamakan *inverse class frequency* (ICF) dan variasinya, *Inverse Class Space density Frequency* ($ICS\delta F$). Dengan ICF dan $ICS\delta F$ ini *term* yang sering muncul pada banyak kelas akan memiliki nilai yang kecil. Metode ini terbukti memiliki

precision dan *recall* yang lebih tinggi daripada TF.IDF dan ICF seperti yang terlihat pada Gambar 2.3 berikut (Fuji Ren dan Mohammad, 2009) :



Gambar 2.3 Class-oriented-indexing TF.IDF, TF.IDF.ICF dan TF.IDF. ICS₆F.

Sumber : Fuji Ren, Mohammad.

2.8 Precision, Recall dan F-Measure

Pengembalian sekumpulan dokumen sebagai jawaban dari *query* pengguna merupakan fungsi dari sistem temu kembali informasi. Untuk mengukur kualitas data *retrieval* digunakan kombinasi *precision* dan *recall*. *Precision* adalah kemampuan sistem untuk mengevaluasi atau menemukan kembali data *top-ranked* yang paling relevan, dan didefinisikan sebagai presentase data yang dikembalikan yang benar-benar relevan terhadap *query* pengguna. *Precision* merupakan proporsi dari suatu set yang diperoleh yang relevan (Maulana, 2014).

Recall mengevaluasi kemampuan sistem temu kembali informasi untuk menemukan semua item yang relevan dari dalam koleksi data dan didefinisikan sebagai presentase data yang relevan terhadap *query* pengguna dan yang diterima. Recall merupakan proporsi dari semua hasil yang relevan di koleksi termasuk hasil yang diperoleh atau dikembalikan (Maulana, 2014).

F-Measure biasa digunakan pada bidang sistem temu kembali informasi untuk mengukur klasifikasi pencarian dokumen dan performa *query classification*. Pada penelitian terdahulu *F-Measure* lebih difokuskan untuk menghitung nilai, namun seiring dengan perkembangan mesin pencari dengan skala besar, kini *F-Measure* lebih menekankan pada kinerja *precision* dan *recall* itu sendiri, sehingga lebih bisa dilihat pada aplikasi secara keseluruhan (Maulana, 2014).

Pada proses perankingan dokumen teks akan dilakukan evaluasi untuk mengukur efektifitas dan relevansi hasil perankingan terhadap *query* yang diinputkan *user*. Proses evaluasi performa efektivitas dari sistem klasifikasi teks

menggunakan standar yang disebut matriks *confusion*. Matriks *confusion* berisi informasi mengenai klasifikasi yang sebenarnya dan prediksi klasifikasi yang dilakukan oleh sistem (Kiftiyani,2014). Berikut ini adalah bentuk tabel matriks *confusion*:

Tabel 2.6 Confusion Matrix

		Actual Class (expectatiom)	
		+	-
Predicted Class (Observation)	+	TP	FP
	-	FN	TN

Sumber: Kiftiyani

Keterangan

TP: *True Positive*, menunjukkan bahwa hasil perankingan oleh sistem memang merupakan dokumen yang sesuai dengan *query*.

FP: *False Positive*, menunjukkan bahwa dokumen yang termasuk dalam hasil perankingan oleh sistem ternyata tidak sesuai dengan *query*.

FN: *False Negative*, menunjukkan bahwa dokumen yang tidak termasuk dalam hasil perankingan oleh sistem ternyata seharusnya sesuai dengan *query*.

TN: *True Negative*, menunjukkan bahwa dokumen yang tidak termasuk hasil perankingan oleh sistem memang sebenarnya tidak sesuai dengan *query*.

Untuk rumus *precision* dapat dilihat dibawah ini (Kiftiyani,2014)

$$\text{precision} = \frac{TP}{TP+FP} \quad (2.10)$$

Nilai *preccission* menunjukkan tingkat ketepatan atau ketelitian dalam pengukuran hasil pengujian atau tingkat ketepatan antara informasi yang diminta oleh *user* dengan hasil jawaban yang diberikan oleh sistem

$$\text{recall} = \frac{TP}{TP+FN} \quad (2.11)$$

Nilai *recall* menunjukkan tingkat jumlah sedikit dan banyaknya kesesuaian informasi yang didapatkan dari hasil percobaan berdasarkan sudut pandang kelas atau label yang digunakan.

$$\text{akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2.12)$$

Semakin tinggi nilai akurasi menunjukkan kesesuaian nilai hasil prediksi pengujian dengan nilai aktual (*ground truth*) yang dibandingkan . *F1 Measure* merupakan bobot *harmonic mean* pada *recall* dan *precision* dengan persamaan sebagai berikut :

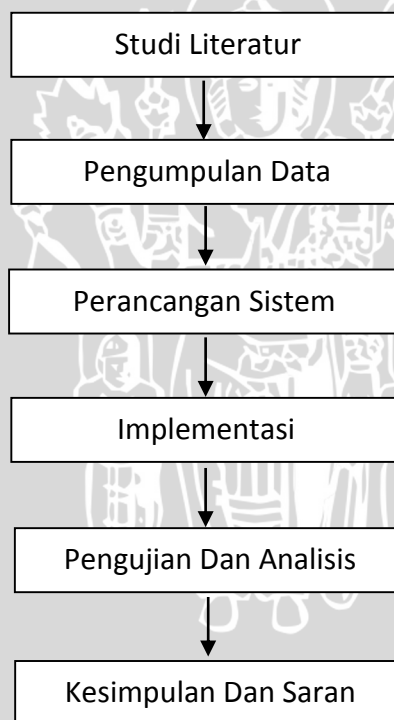
$$F1 = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (2.13)$$

BAB 3 METODOLOGI PENELITIAN

Demi terselesainya penelitian ini, maka dibutuhkan susunan tahapan kegiatan penelitian yang terstruktur dan tepat serta perancangan yang baik. Pada metodologi penelitian ini akan dibahas langkah-langkah dan rancangan yang digunakan dalam pembuatan sistem temu kembali informasi dokumen ide kreatif berbahasa indonesia menggunakan seleksi fitur dengan *metode inverse class space density frequency*.

3.1 Tahapan Penelitian

Bab ini membahas tentang metodologi yang digunakan dalam penelitian ini dan menjelaskan rancangan sistem yang dikembangkan. Metodologi penelitian yang dilakukan dalam penelitian ini yaitu studi literatur, pengumpulan data, analisis kebutuhan, perancangan sistem, implementasi, pengujian dan analisis, dan kesimpulan. Tahapan-tahapan dalam penelitian tersebut dapat diilustrasikan dengan diagram blok metodologi penelitian seperti pada Gambar 3.1 di bawah ini:



Gambar 3.1 Diagram Blok Metodologi Penelitian

3.2 Studi Literatur

Studi Literatur berisikan tentang apa saja sumber yang bisa dijadikan sebagai dasar ilmu yang terkait dengan penelitian ini yaitu "Sistem Temu Kembali Informasi Dokumen Ide Kreatif Berbahasa Indonesia Menggunakan Seleksi Fitur Dengan Metode *Inverse Class Space Density Frequency* (ICS₆F)". Literatur yang digunakan diambil dari buku, internet dan jurnal terkait dengan

1. Sistem Temu Kembali Informasi
2. *Vector Space Model*
3. *Cosine Similarity*
4. *Inverse Class Space Density Frequency*
5. Struktur Morfologi Bahasa Indonesia
6. *Precision, Recall, F-Measure*

3.3 Analisis Kebutuhan

Sistem ini dibuat menggunakan bahasa pemrograman PHP dan SQL *server* sebagai tempat penyimpanan data ide kreatif. Data yang digunakan untuk sistem ini adalah data ide kreatif PJB Unit Paiton pada tahun 2014.

Dalam pembuatan sistem temu kembali informasi ide kreatif ini penerapan memerlukan kebutuhan, baik kebutuhan perangkat keras maupun kebutuhan perangkat lunak serta kebutuhan data yang diperlukan dalam penelitian ini. Kebutuhan tersebut meliputi

1. Kebutuhan Hardware
 - Komputer PC
2. Kebutuhan Software
 - Microsoft Office 2013 sebagai perhitungan manualnya
 - Sublime sebagai *text editor* pembuatan *source code* program
 - XAMPP sebagai penghubung antara sublime dan MySQL
 - MySQL sebagai tempat penyimpanan *database*.
3. Data yang dibutuhkan

Data yang dibutuhkan dalam sistem ini yaitu Data Ide kreatif PJB UP Paiton tahun 2014.

3.4 Perancangan

Perancangan membahas tentang apa yang harus dilakukan untuk pembuatan sistem setelah analisis kebutuhan. Proses perancangan terdiri dari perancangan sistem temu kembali informasi untuk sistem ini sendiri, perancangan *preprocessing*, perancangan metode *inverse class space density frequency*, perancangan cosine similarity dan yang terakhir perancangan perhitungan untuk akurasi data. Hasil untuk proses perancangan ini dibuat agar sistem temu kembali informasi dokumen ide kreatif berbahasa Indonesia menggunakan seleksi fitur dengan metode *inverse class space density frequency* berjalan sebagaimana mestinya.

3.5 Implementasi

Implementasi terdiri dari dua bagian yaitu uraian tentang lingkungan sistem dan uraian tentang implementasi sistem.

3.6 Pengujian

Tahap pengujian digunakan untuk menguji hasil kerja sistem sesuai apa tidak dengan yang diinginkan. Jika tidak sesuai maka dapat dilakukan evaluasi terhadap sistem. Di dalam pengujian ini dilakukan untuk membandingkan besar manakah hasil antara nilai *precision*, *recall* dan *f-measure* untuk data yang menggunakan metode pembobotan lain misalnya pembobotan dokumen bahasa Indonesia dengan TF.IDF.ICF

3.7 Kesimpulan dan Saran

Kesimpulan digunakan sebagai laporan akhir bagaimana kinerja sistem. Sedangkan saran digunakan untuk membantu membangun pengembangan sistem untuk ke depannya.

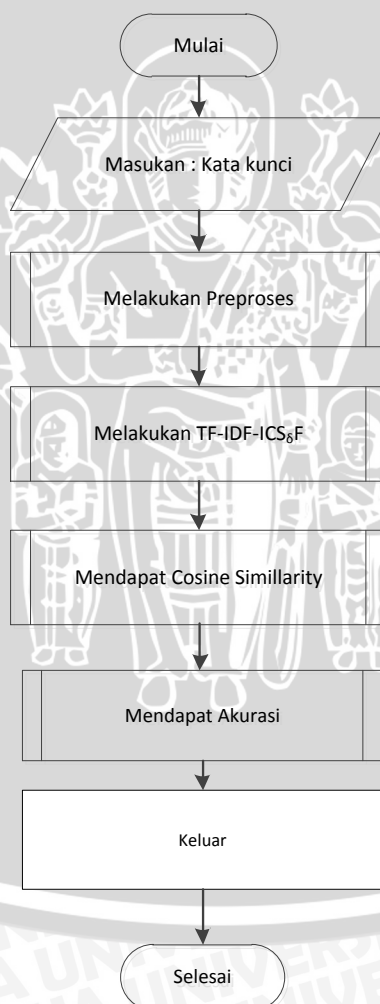


BAB 4 ANALISIS DAN PERANCANGAN

Pada bab ini akan dijelaskan bagaimana gambaran program yang akan dibuat sesuai dengan bahasan yang ada pada bab pendahuluan. Dijelaskan tentang bagaimana mencari pembobotan dokumen ide kreatif berbahasa Indonesia sesuai dengan kata kunci menggunakan *metode inverse class space density frequency*. Setelah dihitung menggunakan metode *TF.IDF.ICSD_F* selanjutnya dihitung dengan *cosine similarity* untuk dihitung jarak kemiripan antar dokumen sebelum akhirnya dihitung keakurasiannya.

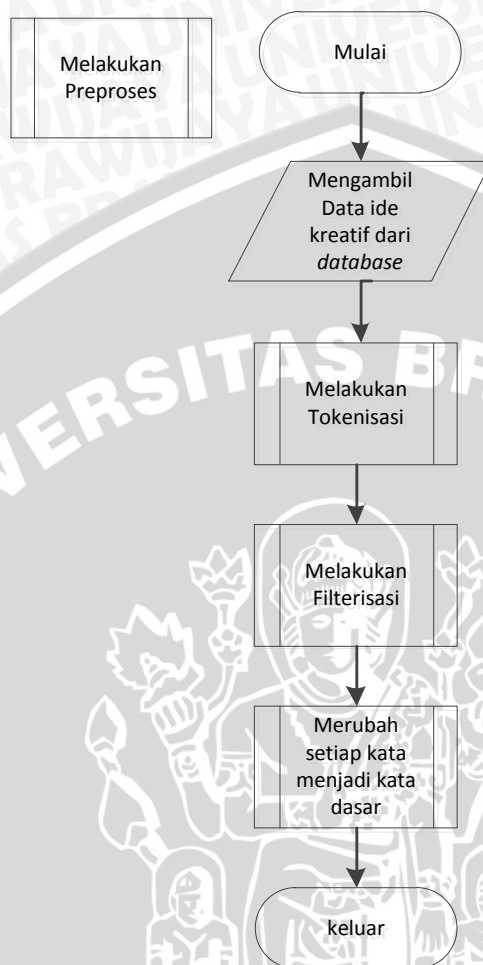
4.1 Teknik Perancangan

Pada Penelitian ini menggunakan teknik Sistem Temu Kembali Informasi dasar. Bagaimana keseluruhan sistem ini berjalan. Untuk penjelasan singkat *flowchart* sistem temu kembali sistem bisa dilihat pada Gambar 4.1 di bawah ini



Gambar 4.1 *Flowchart* sistem temu kembali sistem

Untuk tahap selanjutnya yaitu tahap *preprocessing* terlebih dahulu yang terdiri dari *tokenizing*, *filtering* dan *stemming*. Untuk penjelasan singkat *flowchart preprocessing* bisa dilihat pada Gambar 4.2 dibawah ini :



Gambar 4.2 Flowchart Preprocessing

4.1.1 Flowchart Tokenizing

Tokenizing yaitu pemotongan kata dalam dokumen, lalu mengubah huruf a sampai z yang besar menjadi kecil. Lalu karakter selain huruf dihilangkan. Kemudian dipisahkan berdasarkan tanda pisah. Untuk pemisah pada penelitian ini menggunakan *semicolon* (;) dan spasi. Proses pemisahan kalimat menjadi kata-kata untuk *Tokenizing* bisa dilihat pada Gambar 4.3 *flowchart Tokenizing* berikut ini :



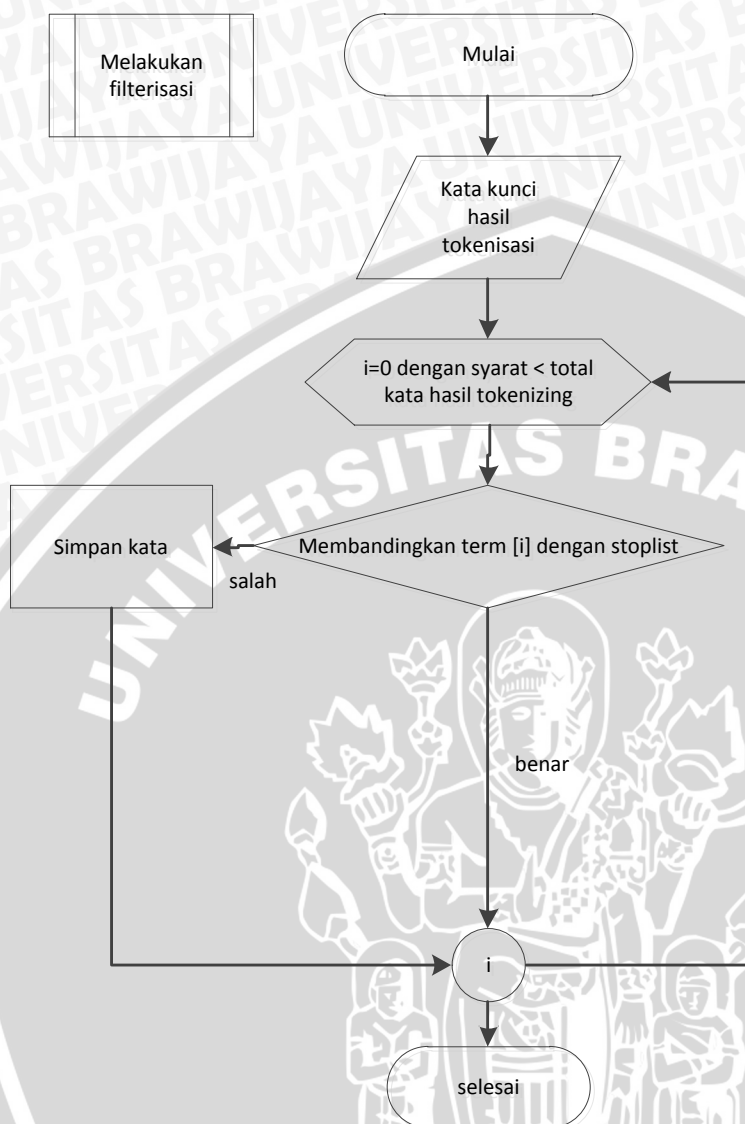
Gambar 4.3 Flowchart Tokenizing

Berikut ini merupakan hasil jika *flowchart* tersebut dijalankan dengan program :

1. *Input* : Modifikasi *wiring penthouse heater* dan *blower* serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.
2. *Output* : modifikasi; *wiring; penthouse; heater; dan; blower; serta; mengintegrasikannya; dengan; modul; pada; epic; III; yang; sudah; ada;*

4.1.2 Flowchart Filtering

Pada proses ini jika ada kata penghubung atau *stoplist* seperti yang, di, dari maka akan dihapus. Diagram proses *filtering* bisa dilihat pada Gambar 4.4 *flowchart filtering* berikut ini



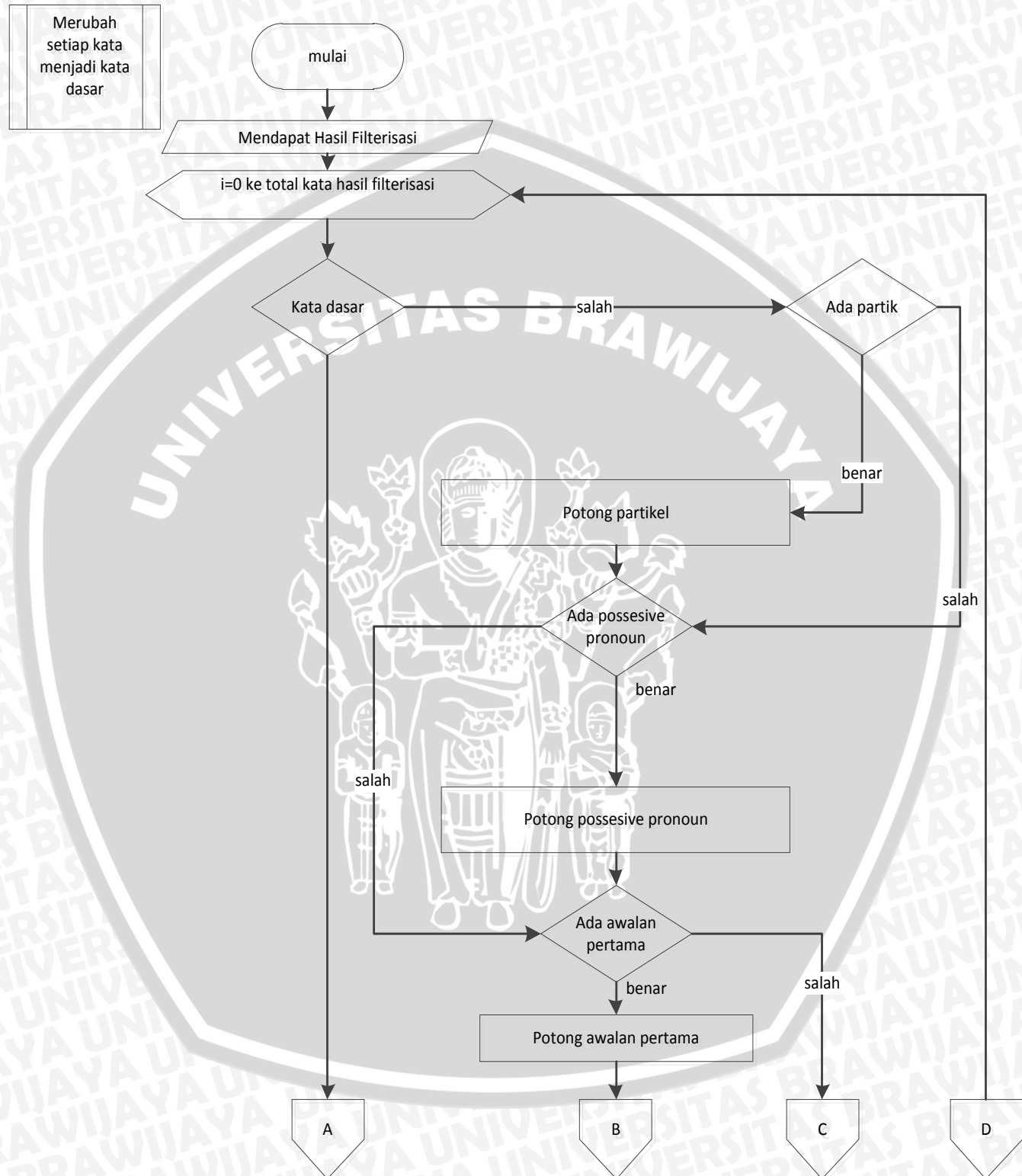
Gambar 4.4 Flowchart Filtering

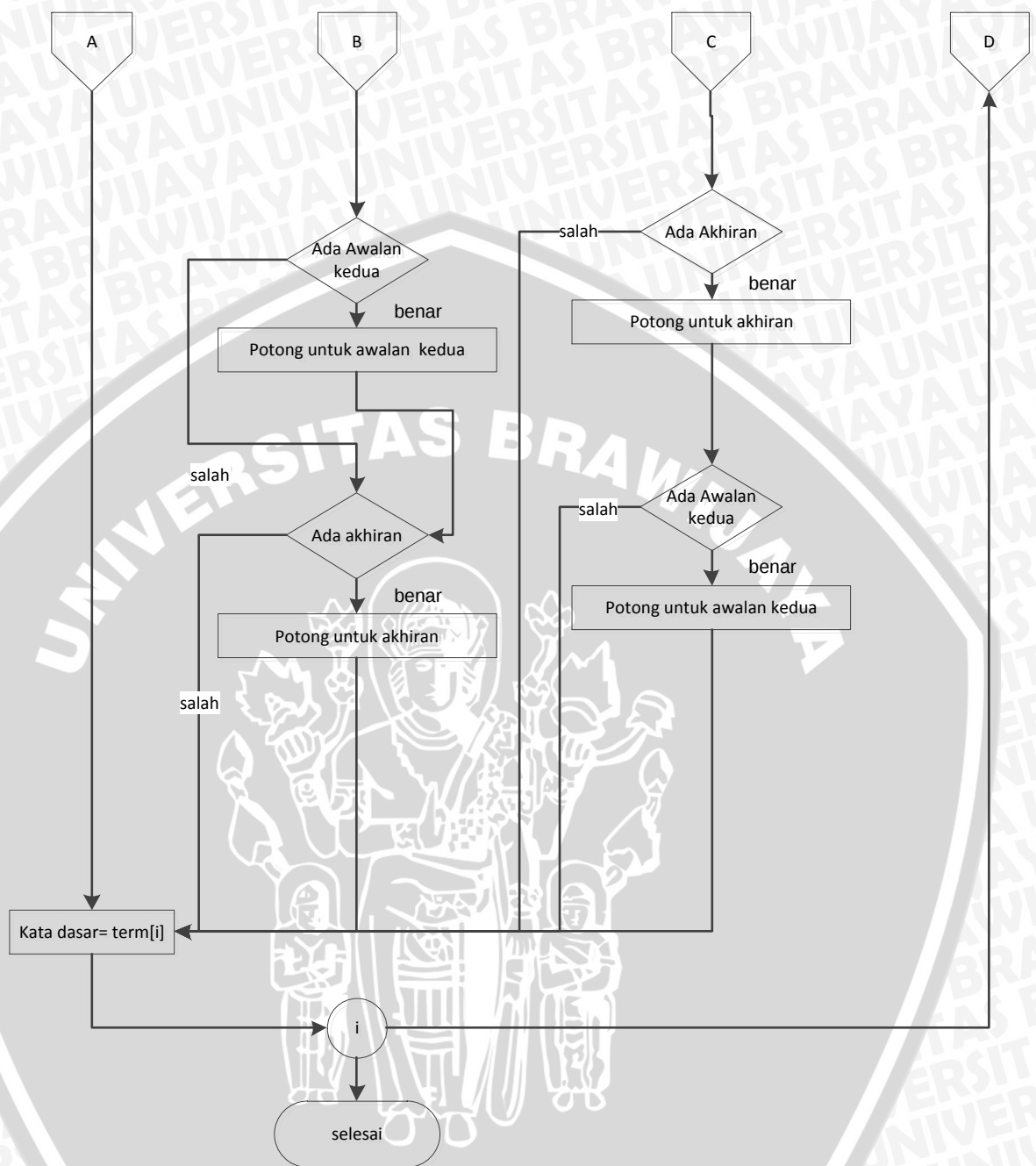
Berikut ini merupakan hasil jika *flowchart* tersebut dijalankan dengan program

1. *Input* (hasil *tokenizing*) : modifikasi; wiring; penthouse; heater; dan; blower; serta; mengintegrasikannya; dengan; modul; pada; epic; III; yang; sudah; ada
2. *Output* : modifikasi; wiring; penthouse; heater; blower; mengintegrasikannya; modul; epic; III

4.1.3 Flowchart Stemming

Sesuai metode Arifin dan Setiono proses penghilangan imbuhan kalimat di awal dan di akhir kalimat dari hasil *filtering* sehingga menjadi kata dasar bisa dilihat pada Gambar 4.5 *Flowchart Stemming* berikut ini





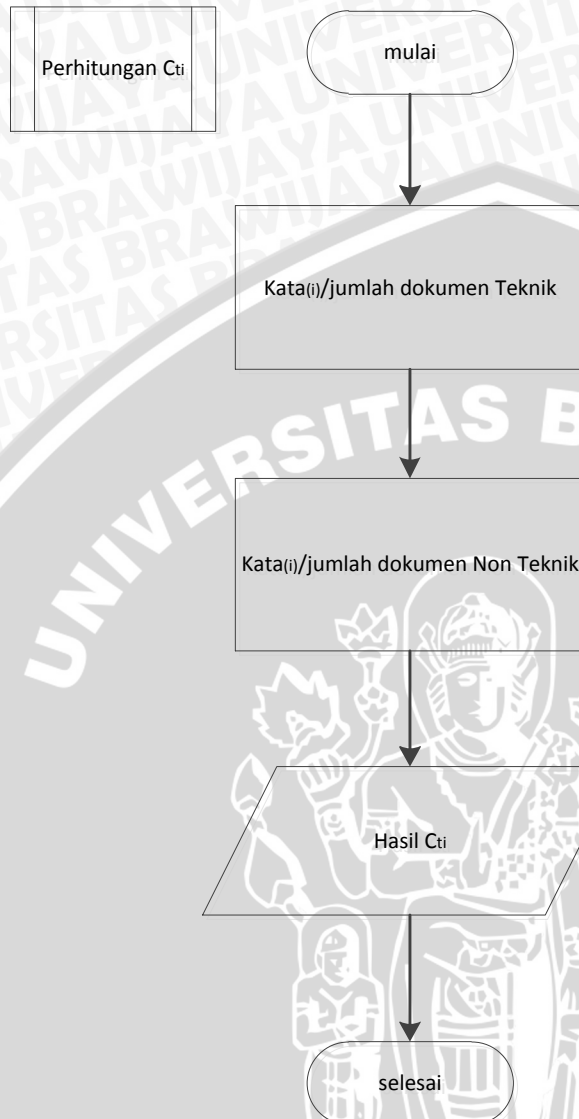
Gambar 4.5 Flowchart Stemming

Berikut ini merupakan hasil jika *flowchart* tersebut dijalankan dengan program

1. *Input* (hasil *filtering*) : modifikasi; wiring; penthouse; heater; blower; mengintegrasikannya; modul; epic; III
2. *Output* : modifikasi; wiring; thouse; heater; blower; integrasi; modul; epic; III

4.1.4 Flowchart Cti

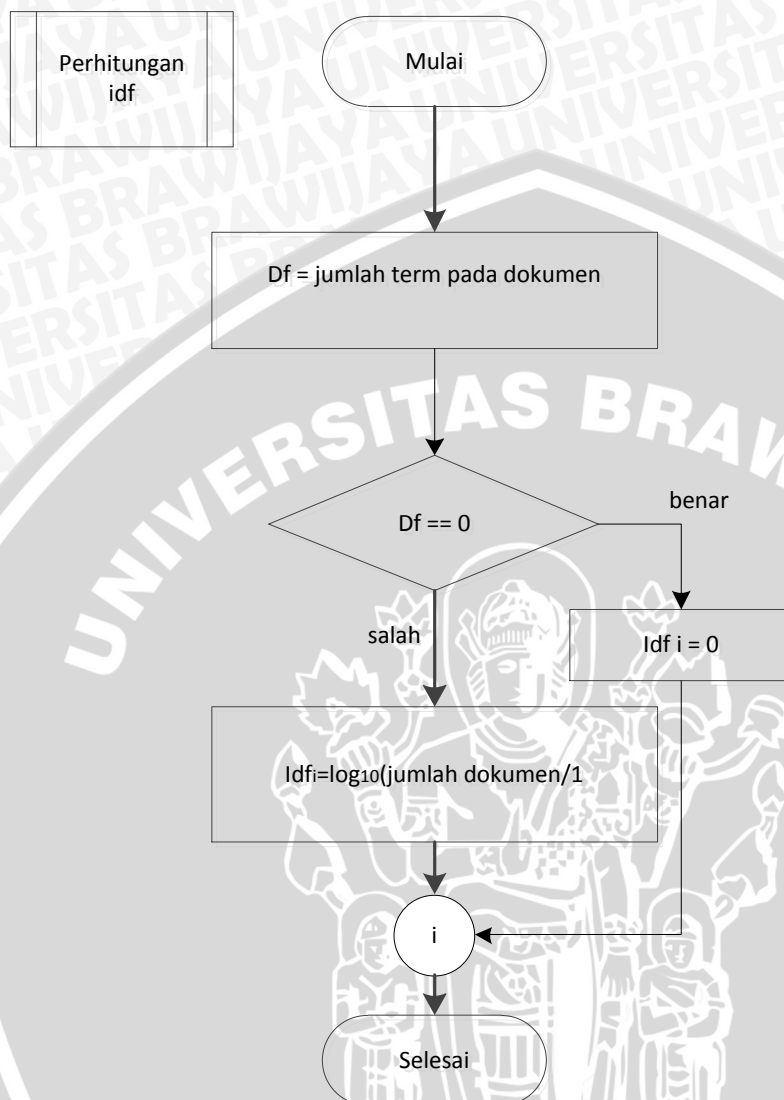
Gambar 4.6 menunjukkan *flowchart* perhitungan *Class Term* ke *i* yang digunakan untuk menghitung nilai *term* pada jenis kelas Teknik dan Non Teknik.



Gambar 4.6 Flowchart Cti

4.1.5 Flowchart Idf

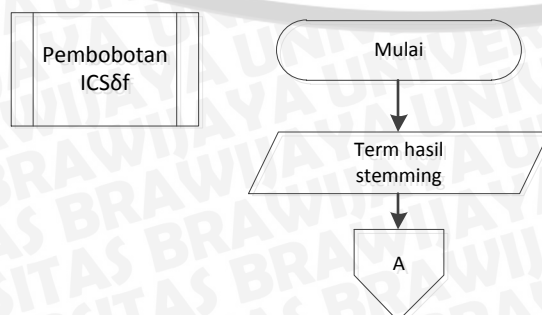
Gambar 4.7 menunjukkan *flowchart* Idf. *TF* yang digunakan adalah *term* yang sudah melalui tahap *stemming*.

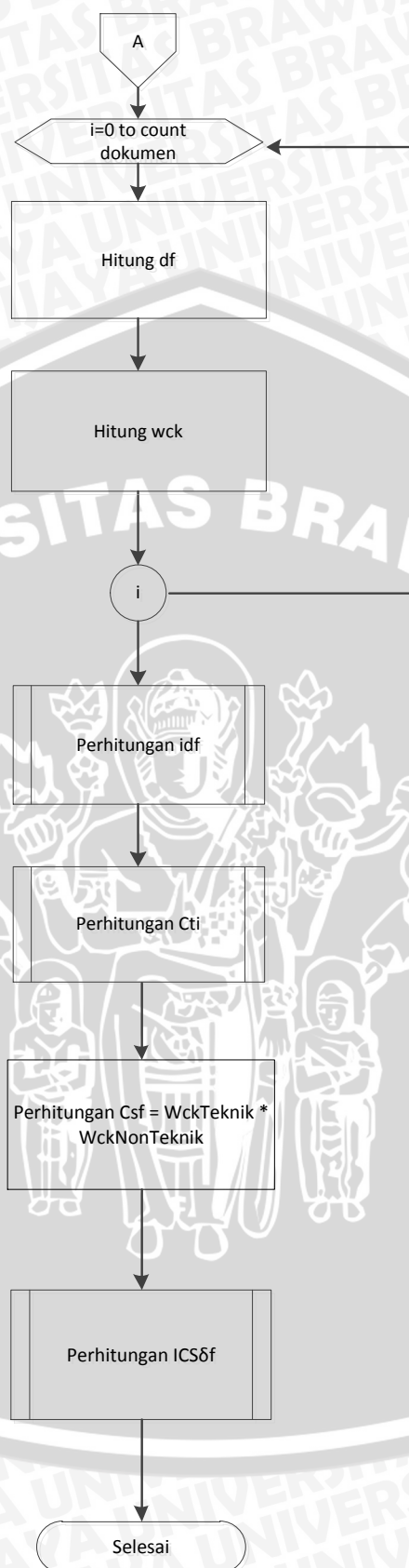


Gambar 4.7 Flowchart Idf

4.1.6 Flowchart Inverse Space Density Space Frequency

Inverse Class Space Density Frequency digunakan untuk pembobotan dokumen yang mengandung kata kunci. Metode ini menggunakan data hasil dari proses preprocessing serta perhitungan TF.IDF. Untuk *penjelasan Flowchart Inverse Space Density Space Frequency* bisa dilihat pada Gambar 4.8 berikut ini

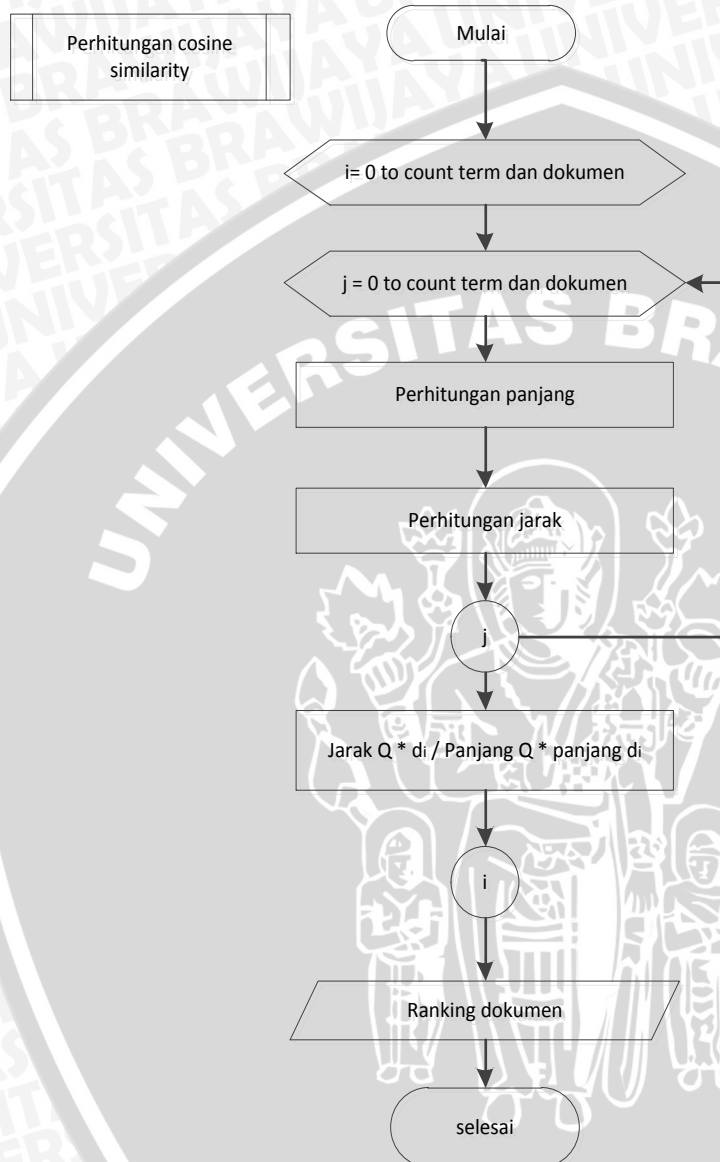




Gambar 4.8 Flowchart Inverse Space Density Space Frequency

4.1.7 Flowchart cosine similarity

Gambar 4.9 menunjukkan *flowchart* perangkat lunak menggunakan *cosine similarity* dengan bobot *term* yang sudah dihitung dengan TF.IDF.



Gambar 4.9 Flowchart cosine similarity

4.2 Perhitungan Halaman Manualisasi

Bagian ini akan dilakukan perhitungan manual dari pemrosesan teks. Dari 207 data Ide Kreatif tahun 2014, diambil 5 dokumen yang terdiri dari kelas Teknik dan Non Teknik. Untuk Tabel 4.1 Dokumen bisa dilihat di bawah ini

Tabel 4.1 Dokumen

Kode	Kelas	Dokumen
D008	NT	memperbanyak tempat biopori disepanjang area unit PLTU #12, memperbanyak penghijauan baik di atas sumber klontong serta penyerapan air hujan disekitar area #12.
D019	T	Pembuatan system instalasi pemanfaatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse, air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paiton dan menerapkan proses recycle, air pada saat step rinse disaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.
D114	NT	PLTU Paiton dapat menerapkan cofiring, antara batu bara dan biodiesel dari kemiri sunan ini. PJB dapat bekerjasama dengan masyarakat sekitar maupun PERHUTANI sebagai supplier bahan mentah, sedangkan PJB mendirikan instalasi pengolahannya. Berikut perbandingan kandungan kemiri sunan dengan biodiesel lainnya. keuntungan cofiring tersebut adalah : peningkatan nilai kalor bahan bakar, sehingga efisiensi pembangkitan meningkat, menunjang pencapaian proper, green power plant karena menggunakan bahan bakar terbarukan.
D147	NT	Mengaktifkan kembali lomba PMK/damkar anatr bagian di intern unit PLTU paiton #12
D161	NT	Saat hari jumat diluangkan 1 jam setelah senam untuk mengenalkan sekaligus untuk membersihkan mesin2 pembangkit tujuannya agar semua karyawan paham atau mengenal mesin2 pembangkit agar semua merasa memiliki terhadap PLTU Paiton yang pada ujungnya rasa tanggung jawab lebih meningkat

4.3 Preprocessing

4.3.1 Tokenizing

Tokenizing adalah proses pemisahan kalimat menjadi per kata dengan pemisah tanda baca “;”. Hasil dari *tokenizing* tercantum pada Tabel 4.2 seperti di bawah ini.

Tabel 4.2 Data hasil *tokenizing*

ID	Kata
D008	memperbanyak; tempat; biopori; disepanjang; area; unit; pltu; memperbanyak; penghijauan; baik; di; atas; sumber; klontong; serta; penyerapan; air; hujan; disekitar; area;
D019	pembuatan; system; instalasi; pemanfaatan; air; sisa; backwash; gravity; sand; filter; dengan; tujuan; menerapkan; proses; reuse; air; bekas; step; water; in; backwash; gsf; digunakan; untuk; penyiraman; tanaman; di; area; pltu; paiton; dan; menerapkan; proses; recycle; air; pada; saat; step; rinse; disaring; kembali; ke; gsf; untuk; digunakan; sebagai; bahan; baku; air; demin;

D114	pltu; paiton; dapat; menerapkan; cofiring; antara; batu; bara; dan; biodiesel; dari; kemiri; sunan; ini; pjb; dapat; bekerjasama; dengan; masyarakat; sekitar; maupun; perhutani; sebagai; supplier; bahan; mentah; sedangkan; pjb; mendirikan; instalasi; pengolahannya; berikut; perbandingan; kandungan; kemiri; sunan; dengan; biodiesel; lainnya; keuntungan; cofiring; tersebut; adalah; peningkatan; nilai; kalor; bahan; bakar; sehingga; efisiensi; pembangkitan; meningkat; menunjang; pencapaian; proper; green; power; plant; karena; menggunakan; bahan; bakar; terbarukan;
D147	mengaktifkan; kembali; lomba; pmk; damkar; anatr; bagian; di; intern; unit; pltu; paiton;
D161	saat; hari; jumat; diluahkan; jam; setelah; senam; untuk; mengenalkan; sekaligus; untuk; membersihkan; mesin; pembangkit; tujuannya; agar; semua; karyawan; paham; atau; mengenal; mesin; pembangkit; agar; semua; merasa; memiliki; terhadap; pltu; paiton; yang; pada; ujungnya; rasa; tanggung; jawab; lebih; meningkat

4.3.2 Filtering

Filtering adalah proses penghapusan kata yang umum dipakai, misal kata penghubung dan lain-lain. Salah satu kata yang termasuk dalam *stoplist* atau kata yang dihilangkan adalah kata “dan” sesuai data yang terlampir. Hasil *filtering* sebagai berikut.

Tabel 4.3 Data hasil *filtering*

ID	Kata
D008	memperbanyak; biopori; disepanjang; area; unit; pltu; memperbanyak; penghijauan; sumber; klontong; penyerapan; air; hujan; disekitar; area;
D019	pembuatan; system; instalasi; pemanfaatan; air; sisa; backwash; gravity; sand; filter; tujuan; menerapkan; proses; reuse; air; bekas; step; water; in; backwash; gsf; penyiraman; tanaman; area; pltu; paiton; menerapkan; proses; recycle; air; step; rinse; disaring; gsf; bahan; baku; air; demin;
D114	pltu; paiton; menerapkan; cofiring; batu; bara; biodiesel; kemiri; sunan; pjb; bekerjasama; masyarakat; perhutani; supplier; bahan; mentah; pjb; mendirikan; instalasi; pengolahannya; perbandingan; kandungan; kemiri; sunan; biodiesel; keuntungan; cofiring; peningkatan; nilai; kalor; bahan; bakar; efisiensi; pembangkitan; meningkat; menunjang; pencapaian; proper; green; power; plant; bahan; bakar; terbarukan;
D147	mengaktifkan; lomba; pmk; damkar; anatr; intern; unit; pltu; paiton;
D161	jumat; diluahkan; jam; senam; mengenalkan; membersihkan; mesin; pembangkit; tujuannya; karyawan; paham; mengenal; mesin; pembangkit; memiliki; pltu; paiton; ujungnya; tanggung; meningkat

4.3.3 Stemming

Stemming adalah proses pemisahan imbuhan pada kata yang telah diolah. Seperti pada contoh kata “latihan” akan menjadi “latih” sesuai metode Arifin dan Setiono. Hasil dari *stemming* tercantum pada Tabel 4.4 seperti dibawah ini:

Tabel 4.4 Data hasil *stemming*

ID	Kata
D008	perbanyak; biopor; sepanjang; area; unit; pltu; perbanyak; hijau; sumber; klontong; serap; air; hujan; sekitar; area;
D019	buat; system; instalasi; anfaat; air; sisa; backwash; gravity; sand; filter; tuju; terap; proses; reuse; air; bekas; step; water; in; backwash; gsf; siram; tanam; area; pltu; paiton; terap; proses; recycle; air; step; rinse; saring; gsf; bahan; baku; air; demin;
D114	pltu; paiton; terap; cofiring; batu; bara; biodiesel; kemiri; sunan; pjb; kerjasama; masyarakat; hutan; supplier; bahan; mentah; pjb; diri; instalasi; olah; banding; kandung; kemiri; sunan; biodiesel; untung; confiring; ingkat; nilai; kalor; bahan; bakar; efisiensi; bangkit; tingkat; tunjang; capai; proper; green; power; plant; bahan; bakar; baru;
D147	aktif; lomba; pmk; damkar; anatr; intern; unit; pltu; paiton;
D161	jumat; luang; jam; senam; enal; sih; mesin; bangkit; tuju; karyawan; paham; enal; mesin; bangkit; pilik; pltu; paiton; ujung; tanggung; tingkat

4.3.4 TF.IDF.ICS_δF

Algoritma *TF.IDF.ICS_δF* digunakan untuk pembobotan kata. Untuk 5 data tersebut saat diperhitungkan dengan metode *TF.IDF.ICS_δF* maka diperoleh hasil seperti Tabel 4.5 berikut ini

Tabel 4.5 Perhitungan *TF.IDF.ICS_δF* 1

Term	Q	TF					Df	Wck		Cti		Csf	idf	ics _δ f
		D8	D19	D114	D147	D161		T	NT	T	NT			
perbanyak	0	2	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903
biopor	0	1	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903
sepanjang	0	1	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903
area	0	2	1	0	0	0	2	1	1	1	0,25	1,25	0,397	0,204
unit	0	1	0	0	1	0	2	0	1	0	0,25	0,25	0,397	0,903
PLTU	0	1	1	1	1	1	5	1	1	1	0,25	1,25	0	0,204
hijau	0	1	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903
sumber	0	1	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903

klontong	0	1	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903
serap	0	1	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903
air	1	1	4	0	0	0	2	1	1	1	0,25	1,25	0,397	0,204
hujan	0	1	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903
sekitar	0	1	0	0	0	0	1	0	1	0	0,25	0,25	0,698	0,903
buat	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
system	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
instalasi	0	0	1	1	0	0	2	1	1	1	0,25	1,25	0,397	0,204
anfaat	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
sisa	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
backwash	0	0	2	0	0	0	1	1	0	1	0	1	0,698	0,301
gravity	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
sand	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
filter	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
tuju	0	0	1	0	0	1	2	1	1	1	0,25	1,25	0,397	0,204
terap	0	0	2	1	0	0	2	1	1	1	0,25	1,25	0,397	0,204
proses	0	0	2	0	0	0	1	1	0	1	0	1	0,698	0,301
reuse	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
bekas	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
step	0	0	2	0	0	0	1	1	0	1	0	1	0,698	0,301
water	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
in	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
GSF	0	0	2	0	0	0	1	1	0	1	0	1	0,698	0,301
siram	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
tanam	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
Paiton	0	0	1	1	1	1	4	1	1	1	0,25	1,25	0,096	0,204
recycle	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
rinse	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
saring	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
bahan	0	0	1	3	0	0	2	1	1	1	0,25	1,25	0,397	0,204
baku	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301

demin	0	0	1	0	0	0	1	1	0	1	0	1	0,698	0,301
cofiring	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
batu	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
bara	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
biodiesel	0	0	0	2	0	0	1	0	1	0	0,25	0,25	0,698	0,903
kemiri	0	0	0	2	0	0	1	0	1	0	0,25	0,25	0,698	0,903
sunan	0	0	0	2	0	0	1	0	1	0	0,25	0,25	0,698	0,903
PJB	0	0	0	2	0	0	1	0	1	0	0,25	0,25	0,698	0,903
kerjasama	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
masyarakat	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
PERHUTANI	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
supplier	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
mentah	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
diri	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
olah	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
ikut	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
banding	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
kandung	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
untung	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
confiring	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
ingkat	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
nilai	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
kalor	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
bakar	0	0	0	2	0	0	1	0	1	0	0,25	0,25	0,698	0,903
efisiensi	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
bangkit	0	0	0	1	0	2	2	0	1	0	0,25	0,25	0,397	0,903
tingkat	0	0	0	1	0	1	2	0	1	0	0,25	0,25	0,397	0,903
tunjang	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
capai	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
proper	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
green	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903

power	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
plant	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
baru	0	0	0	1	0	0	1	0	1	0	0,25	0,25	0,698	0,903
aktif	0	0	0	0	1	0	1	0	1	0	0,25	0,25	0,698	0,903
lomba	0	0	0	0	1	0	1	0	1	0	0,25	0,25	0,698	0,903
PMK	0	0	0	0	1	0	1	0	1	0	0,25	0,25	0,698	0,903
damkar	0	0	0	0	1	0	1	0	1	0	0,25	0,25	0,698	0,903
anatr	0	0	0	0	1	0	1	0	1	0	0,25	0,25	0,698	0,903
Intern	0	0	0	0	1	0	1	0	1	0	0,25	0,25	0,698	0,903
Paiton	0	0	1	1	1	1	4	1	1	1	0,25	1,25	0,096	0,204
Saat	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Jumat	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Luang	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Jam	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Senam	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Enal	0	0	0	0	0	2	1	0	1	0	0,25	0,25	0,698	0,903
Sih	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Mesin	0	0	0	0	0	2	1	0	1	0	0,25	0,25	0,698	0,903
karyawan	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Paham	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Pilik	0	0	0	0	0	1	1	0	1	0	0,25	0,25	0,698	0,903
Ujung	0	0	0	0	0	1	1	0	1	0	0,25	0	0,698	0,903
Tanggung	0	0	0	0	0	1	1	0	1	0	0,25	0	0,698	0,903

Nilai di atas didapat dari :

1. *Term* pada tabel diatas merupakan 5 dokumen teknik dan non teknik yang sudah melalui proses *preprocessing*.
2. TF atau *Term Frequency* merupakan seberapa banyak kata muncul pada tiap dokumen. Karena pada "air" pada dokumen 8 muncul sebanyak 1 maka TF untuk dokumen 1 adalah 1.

3. df atau *Degree of freedom* merupakan jumlah banyaknya tiap term pada seluruh dokumen. *Degree of freedom* sejumlah 2, untuk term "air" muncul pada dokumen 8 dan dokumen 19
4. Wck merupakan jumlah *query* untuk masing-masing kelas Teknik dan Non teknik. Untuk kata "air" pada kelas Teknik dan NonTeknik bernilai 1 karena kata tersebut ada pada kedua kelas.
5. Cti adalah jumlah *term* untuk dokumen kelas Teknik dan Non Teknik. Misalkan "air" untuk kelas Teknik adalah 1 maka hasil nya adalah 1 dibagi dengan banyak kelas Teknik yaitu 1 maka diperoleh nilai 1 . Sedangkan "air" untuk kelas non teknik ada 1 . Sehingga 1 dibagi dengan banyak kelas non teknik yaitu 4 maka hasilnya 0,25. Contoh perhitungan pembobotan kata untuk term "air" pada kelas teknik menggunakan persamaan (2.6)

$$C(\text{modifikasi}) = \sum_{ck} CS\delta(\text{modifikasi}) = 0,25$$

6. Csf merupakan jumlah antara kelas teknik dan nonteknik pada Cti untuk masing-masing *term*.
7. Idf merupakan hasil dari log10 dari jumlah dokumen dibagi dengan hasil df tiap term. Jadi pada tabel untuk term "air" didapatkan idf sejumlah 0,397 dari $\log(5/2) \cdot 5$ merupakan jumlah keseluruhan dokumen. Untuk perhitungannya bisa dilihat pada persamaan 2.2 berikut ini

$$Idf_{(\text{modifikasi})} = \log\left(\frac{n}{df_{\text{air}}}\right)$$

$$Idf_{(\text{modifikasi})} = \log\left(\frac{5}{2}\right)$$

$$Idf_{(\text{modifikasi})} = 0,397$$

8. ICS_δF merupakan log untuk jumlah jenis kelas yang ada yaitu Teknik dan Non teknik sebanyak 2 lalu dibagi dengan hasil dari Csf untuk tiap *term*. Untuk perhitungannya bisa dilihat pada persamaan 2.7 berikut ini

$$ICS\delta F(\text{air}) = \log\left(\frac{2}{CS\delta}\right)$$

$$ICS\delta F(\text{air}) = \log\left(\frac{2}{1,25}\right)$$

$$ICS\delta F(\text{air}) = 0,204$$

4.3.5 Cosine Similarity

Algoritma *Cosine Similarity* yang digunakan untuk mencari solusi kemiripan antar dokumen dapat dilihat dibawah Tabel 4.6 ,4.7 di bawah ini

Tabel 4.6 Perhitungan $TF.IDF.ICSD F 2$

Term	TF				idf	icsdf	Weight			
	Q	D8	D19	D114			Q	D8	D19	D114
perbanyak	0	2	0	0	0,698	0,903	0	1,262	0	0
biopor	0	1	0	0	0,698	0,903	0	0,631	0	0
sepanjang	0	1	0	0	0,698	0,903	0	0,631	0	0
area	0	2	1	0	0,397	0,204	0	0,16	0,081	0
Unit	0	1	0	0	0,397	0,903	0	0,359	0	0
PLTU	0	1	1	1	0	0,204	0	0	0	0
hijau	0	1	0	0	0,698	0,903	0	0,631	0	0
sumber	0	1	0	0	0,698	0,903	0	0,631	0	0
klontong	0	1	0	0	0,698	0,903	0	0,631	0	0
serap	0	1	0	0	0,698	0,903	0	0,631	0	0
air	1	1	4	0	0,397	0,204	0,081	0,081	0,324	0
hujan	0	1	0	0	0,698	0,903	0	0,631	0	0
sekitar	0	1	0	0	0,698	0,903	0	0,631	0	0
buat	0	0	1	0	0,698	0,301	0	0	0,210	0
system	0	0	1	0	0,698	0,301	0	0	0,210	0
instalasi	0	0	1	1	0,397	0,204	0	0	0,081	0,081
anfaat	0	0	1	0	0,698	0,301	0	0	0,210	0
sisa	0	0	1	0	0,698	0,301	0	0	0,210	0
backwash	0	0	2	0	0,698	0,301	0	0	0,420	0
gravity	0	0	1	0	0,698	0,301	0	0	0,210	0
sand	0	0	1	0	0,698	0,301	0	0	0,210	0
filter	0	0	1	0	0,698	0,301	0	0	0,210	0
tuju	0	0	1	0	0,397	0,204	0	0	0,081	0
terap	0	0	2	1	0,397	0,204	0	0	0,162	0,081
proses	0	0	2	0	0,698	0,301	0	0	0,420	0
reuse	0	0	1	0	0,698	0,301	0	0	0,210	0
bekas	0	0	1	0	0,698	0,301	0	0	0,210	0

step	0	0	2	0	0,698	0,301	0	0	0,420	0
water	0	0	1	0	0,698	0,301	0	0	0,210	0
in	0	0	1	0	0,698	0,301	0	0	0,210	0
GSF	0	0	2	0	0,698	0,301	0	0	0,420	0
siram	0	0	1	0	0,698	0,301	0	0	0,210	0
tanam	0	0	1	0	0,698	0,301	0	0	0,210	0
Paiton	0	0	1	1	0,096	0,204	0	0	0,019	0,019
recycle	0	0	1	0	0,698	0,301	0	0	0,210	0
rinse	0	0	1	0	0,698	0,301	0	0	0,210	0
saring	0	0	1	0	0,698	0,301	0	0	0,210	0
bahan	0	0	1	3	0,397	0,204	0	0	0,081	0,243
baku	0	0	1	0	0,698	0,301	0	0	0,210	0
demin	0	0	1	0	0,698	0,301	0	0	0,210	0
cofiring	0	0	0	1	0,698	0,903	0	0	0	0,631
batu	0	0	0	1	0,698	0,903	0	0	0	0,631
bara	0	0	0	1	0,698	0,903	0	0	0	0,631
biodiesel	0	0	0	2	0,698	0,903	0	0	0	1,262
kemiri	0	0	0	2	0,698	0,903	0	0	0	1,262
sunan	0	0	0	2	0,698	0,903	0	0	0	1,262
PJB	0	0	0	2	0,698	0,903	0	0	0	1,262
kerjasama	0	0	0	1	0,698	0,903	0	0	0	0,631
masyarakat	0	0	0	1	0,698	0,903	0	0	0	0,631
PERHUTANI	0	0	0	1	0,698	0,903	0	0	0	0,631
supplier	0	0	0	1	0,698	0,903	0	0	0	0,631
mentah	0	0	0	1	0,698	0,903	0	0	0	0,631
diri	0	0	0	1	0,698	0,903	0	0	0	0,631
olah	0	0	0	1	0,698	0,903	0	0	0	0,631
ikut	0	0	0	1	0,698	0,903	0	0	0	0,631
banding	0	0	0	1	0,698	0,903	0	0	0	0,631
kandung	0	0	0	1	0,698	0,903	0	0	0	0,631
untung	0	0	0	1	0,698	0,903	0	0	0	0,631
confiring	0	0	0	1	0,698	0,903	0	0	0	0,631
ingkat	0	0	0	1	0,698	0,903	0	0	0	0,631

nilai	0	0	0	1	0,698	0,903	0	0	0	0,631
kalor	0	0	0	1	0,698	0,903	0	0	0	0,631
bakar	0	0	0	2	0,698	0,903	0	0	0	1,262
efisiensi	0	0	0	1	0,698	0,903	0	0	0	0,631
bangkit	0	0	0	1	0,397	0,903	0	0	0	0,359
tingkat	0	0	0	1	0,397	0,903	0	0	0	0,359
tunjang	0	0	0	1	0,698	0,903	0	0	0	0,631
capai	0	0	0	1	0,698	0,903	0	0	0	0,631
proper	0	0	0	1	0,698	0,903	0	0	0	0,631
green	0	0	0	1	0,698	0,903	0	0	0	0,631
power	0	0	0	1	0,698	0,903	0	0	0	0,631
plant	0	0	0	1	0,698	0,903	0	0	0	0,631
baru	0	0	0	1	0,698	0,903	0	0	0	0,631
aktif	0	0	0	0	0,698	0,903	0	0	0	0
lomba	0	0	0	0	0,698	0,903	0	0	0	0
PMK	0	0	0	0	0,698	0,903	0	0	0	0
damkar	0	0	0	0	0,698	0,903	0	0	0	0
anatr	0	0	0	0	0,698	0,903	0	0	0	0
intern	0	0	0	0	0,698	0,903	0	0	0	0
paiton	0	0	1	1	0,096	0,204	0	0	0,019	0,019
Saat	0	0	0	0	0,698	0,903	0	0	0	0
jumat	0	0	0	0	0,698	0,903	0	0	0	0
luang	0	0	0	0	0,698	0,903	0	0	0	0
jam	0	0	0	0	0,698	0,903	0	0	0	0
senam	0	0	0	0	0,698	0,903	0	0	0	0
enal	0	0	0	0	0,698	0,903	0	0	0	0
sih	0	0	0	0	0,698	0,903	0	0	0	0
mesin	0	0	0	0	0,698	0,903	0	0	0	0
karyawan	0	0	0	0	0,698	0,903	0	0	0	0
paham	0	0	0	0	0,698	0,903	0	0	0	0
pilik	0	0	0	0	0,698	0,903	0	0	0	0
ujung	0	0	0	0	1	1	0	1	0	0,25
tanggung	0	0	0	0	1	1	0	1	0	0,25

Tabel 4.7 Perhitungan $TF.IDF.ICSF_3$

Term	TF		idf	icsdf	Weight	
	D147	D161			D147	D161
perbanyak	0	0	0,698	0,903	0	0
biopor	0	0	0,698	0,903	0	0
sepanjang	0	0	0,698	0,903	0	0
area	0	0	0,397	0,204	0	0
unit	0	1	0,397	0,903	0,359	0
PLTU	1	1	0	0,204	0	0
hijau	0	0	0,698	0,903	0	0
sumber	0	0	0,698	0,903	0	0
klontong	0	0	0,698	0,903	0	0
serap	0	0	0,698	0,903	0	0
air	0	0	0,397	0,204	0	0
hujan	0	0	0,698	0,903	0	0
sekitar	0	0	0,698	0,903	0	0
buat	0	0	0,698	0,301	0	0
system	0	0	0,698	0,301	0	0
instalasi	1	0	0,397	0,204	0	0
anfaat	0	0	0,698	0,301	0	0
sisia	0	0	0,698	0,301	0	0
backwash	0	0	0,698	0,301	0	0
gravity	0	0	0,698	0,301	0	0
sand	0	0	0,698	0,301	0	0
filter	0	0	0,698	0,301	0	0
tuju	0	0	0,397	0,204	0	0,081
terap	1	0	0,397	0,204	0	0
proses	0	0	0,698	0,301	0	0
reuse	0	0	0,698	0,301	0	0
bekas	0	0	0,698	0,301	0	0
step	0	0	0,698	0,301	0	0
water	0	0	0,698	0,301	0	0

in	0	0	0,698	0,301	0	0
GSF	0	0	0,698	0,301	0	0
siram	0	0	0,698	0,301	0	0
tanam	0	0	0,698	0,301	0	0
Paiton	1	1	0,096	0,204	0,019	0,019
recycle	0	0	0,698	0,301	0	0
rinse	0	0	0,698	0,301	0	0
saring	0	0	0,698	0,301	0	0
bahan	3	0	0,397	0,204	0	0
baku	0	0	0,698	0,301	0	0
demin	0	0	0,698	0,301	0	0
cofiring	1	0	0,698	0,903	0	0
batu	1	0	0,698	0,903	0	0
bara	1	0	0,698	0,903	0	0
biodiesel	2	0	0,698	0,903	0	0
kemiri	2	0	0,698	0,903	0	0
sunan	2	0	0,698	0,903	0	0
PJB	2	0	0,698	0,903	0	0
kerjasama	1	0	0,698	0,903	0	0
masyarakat	1	0	0,698	0,903	0	0
PERHUTANI	1	0	0,698	0,903	0	0
supplier	1	0	0,698	0,903	0	0
mentah	1	0	0,698	0,903	0	0
diri	1	0	0,698	0,903	0	0
olah	1	0	0,698	0,903	0	0
ikut	1	0	0,698	0,903	0	0
banding	1	0	0,698	0,903	0	0
kandung	1	0	0,698	0,903	0	0
untung	1	0	0,698	0,903	0	0
confiring	1	0	0,698	0,903	0	0
ingkat	1	0	0,698	0,903	0	0
nilai	1	0	0,698	0,903	0	0
kalor	1	0	0,698	0,903	0	0

bakar	2	0	0,698	0,903	0	0
efisiensi	1	0	0,698	0,903	0	0
bangkit	1	0	0,397	0,903	0	0,718
tingkat	1	0	0,397	0,903	0	0,359
tunjang	1	0	0,698	0,903	0	0
capai	1	0	0,698	0,903	0	0
proper	1	0	0,698	0,903	0	0
green	1	0	0,698	0,903	0	0
power	1	0	0,698	0,903	0	0
plant	1	0	0,698	0,903	0	0
baru	1	0	0,698	0,903	0	0
aktif	0	1	0,698	0,903	0,631	0
lomba	0	1	0,698	0,903	0,631	0
PMK	0	1	0,698	0,903	0,631	0
damkar	0	1	0,698	0,903	0,631	0
anatr	0	1	0,698	0,903	0,631	0
intern	0	1	0,698	0,903	0,631	0
paiton	1	1	0,096	0,204	0,019	0,019
Saat	0	0	0,698	0,903	0	0,631
jumat	0	0	0,698	0,903	0	0,631
luang	0	0	0,698	0,903	0	0,631
jam	0	0	0,698	0,903	0	0,631
senam	0	0	0,698	0,903	0	0,631
enal	0	0	0,698	0,903	0	1,262
sih	0	0	0,698	0,903	0	0,631
mesin	0	0	0,698	0,903	0	1,262
karyawan	0	0	0,698	0,903	0	0,631
paham	0	0	0,698	0,903	0	0,631
pilik	0	0	0,698	0,903	0	0,631
ujung	0	0	1	1	0	0,698
tanggung	0	0	1	1	0	0,698

Untuk perhitungan weight pada term “air” bisa dilihat pada rumus di bawah ini

$$\text{Weight}_{(ti)} = \text{TF}_{d8(ti)} * \text{Idf}_{(ti)} * \text{ICS}_{\delta f(ti)}$$

$$\text{Weight}_{(air)} = \text{TF}_{d8(air)} * \text{Idf}_{(air)} * \text{ICS}_{\delta f(air)}$$

$$\text{Weight}_{(air)} = 1 * 0,397 * 0,204 = 0,081$$

Tabel 4.8 Perhitungan Cosine Similarity

Term	Weight ²					
	Q	D8	D19	D114	D147	D161
perbanyak	0	1,593	0	0	0	0
biopor	0	0,398	0	0	0	0
sepanjang	0	0,398	0	0	0	0
area	0	0,026	0,006	0	0	0
unit	0	0,129	0	0	0,129	0
PLTU	0	0	0	0	0	0
hijau	0	0,398	0	0	0	0
sumber	0	0,398	0	0	0	0
klontong	0	0,398	0	0	0	0
serap	0	0,398	0	0	0	0
air	0,006	0,006	0,105	0	0	0
hujan	0	0,398	0	0	0	0
sekitar	0	0,398	0	0	0	0
buat	0	0	0,044	0	0	0
system	0	0	0,044	0	0	0
instalasi	0	0	0,006	0,006	0	0
anfaat	0	0	0,044	0	0	0
sisa	0	0	0,044	0	0	0
backwash	0	0	0,177	0	0	0
gravity	0	0	0,044	0	0	0
sand	0	0	0,044	0	0	0
filter	0	0	0,044	0	0	0
tuju	0	0	0,006	0	0	0,006
terap	0	0	0,026	0,006	0	0
proses	0	0	0,177	0	0	0
reuse	0	0	0,044	0	0	0

bekas	0	0	0,044	0	0	0
step	0	0	0,177	0	0	0
water	0	0	0,044	0	0	0
in	0	0	0,044	0	0	0
GSF	0	0	0,177	0	0	0
siram	0	0	0,044	0	0	0
tanam	0	0	0,044	0	0	0
Paiton	0	0	0,0003	0,0003	0,0003	0,0003
recycle	0	0	0,044	0	0	0
rinse	0	0	0,044	0	0	0
saring	0	0	0,044	0	0	0
bahan	0	0	0,006	0,059	0	0
baku	0	0	0,044	0	0	0
demin	0	0	0,044	0	0	0
cofiring	0	0	0	0,398	0	0
batu	0	0	0	0,398	0	0
bara	0	0	0	0,398	0	0
biodiesel	0	0	0	1,593	0	0
kemiri	0	0	0	1,593	0	0
sunan	0	0	0	1,593	0	0
PJB	0	0	0	1,593	0	0
kerjasama	0	0	0	0,398	0	0
masyarakat	0	0	0	0,398	0	0
PERHUTANI	0	0	0	0,398	0	0
supplier	0	0	0	0,398	0	0
mentah	0	0	0	0,398	0	0
diri	0	0	0	0,398	0	0
olah	0	0	0	0,398	0	0
ikut	0	0	0	0,398	0	0
banding	0	0	0	0,398	0	0
kandung	0	0	0	0,398	0	0
untung	0	0	0	0,398	0	0
confiring	0	0	0	0,398	0	0

ingkat	0	0	0	0,398	0	0
nilai	0	0	0	0,398	0	0
kalor	0	0	0	0,398	0	0
bakar	0	0	0	1,593	0	0
efisiensi	0	0	0	0,398	0	0
bangkit	0	0	0	0,129	0	0,516
tingkat	0	0	0	0,129	0	0,129
tunjang	0	0	0	0,398	0	0
capai	0	0	0	0,398	0	0
proper	0	0	0	0,398	0	0
green	0	0	0	0,398	0	0
power	0	0	0	0,398	0	0
plant	0	0	0	0,398	0	0
baru	0	0	0	0,398	0	0
aktif	0	0	0	0	0,398	0
lomba	0	0	0	0	0,398	0
PMK	0	0	0	0	0,398	0
damkar	0	0	0	0	0,398	0
anatr	0	0	0	0	0,398	0
intern	0	0	0	0	0,398	0
paiton	0	0	0,0003	0,0003	0,0003	0,0003
Saat	0	0	0	0	0	0,398
jumat	0	0	0	0	0	0,398
luang	0	0	0	0	0	0,398
jam	0	0	0	0	0	0,398
senam	0	0	0	0	0	0,398
enal	0	0	0	0	0	1,593
sih	0	0	0	0	0	0,398
mesin	0	0	0	0	0	1,593
karyawan	0	0	0	0	0	0,398
paham	0	0	0	0	0	0,398
pilik	0	0	0	0	0	0,398
ujung	0	0	0	0	0	0,398

tanggung	0	0	0	0	0	0,398
----------	---	---	---	---	---	-------

Untuk perhitungan $Weight^2$ pada term “air” bisa dilihat pada rumus di bawah ini

$$(Weight_{(d_i, t_i)})^2 = (Weight_{(d_i, air)})^2$$

$$(Weight_{(d8, air)})^2 = 0,081^2$$

$$(Weight_{(d8, air)})^2 = 0,006$$

Untuk hasil panjang vektor bisa dilihat pada Tabel 4.8 di bawah ini

Tabel 4.8 Panjang Vektor

Panjang Vektor	
Q	0,081228
D8	2,223421
D19	1,290119
D114	4,31979
D147	1,587659

Perhitungan perankingan dokumen untuk *query* “air” diatas dihitung dengan *cosine similarity* menggunakan rumus berikut ini

$$\text{panjang vektor} = \sqrt{\sum_i^t \text{weight}_i^2}$$

Jarak dari *query* “air” untuk kelima dokumen diatas bisa dilihat pada Tabel 4.9 di bawah ini

Tabel 4.9 Jarak Terhadap query

Jarak Terhadap query	
D8	0,006598
D19	0,026392
D114	0
D147	0
D161	0

Hasil tabel diatas diperoleh dari perhitungan rumus dibawah ini :

$$\text{jarak} = \sum_{i=0}^n \text{weight}_q \cdot \text{weight}_{d_i}$$

$$\text{jarak} = 0,006598$$

Perhitungan hasil cosine dengan *query* “air” untuk kelima dokumen bisa dilihat pada Tabel 4.10 Hasil *cosine* di bawah ini

Tabel 4.9 Hasil *cosine*

Hasil <i>Cosine</i> Terhadap Q	
D8	0,036533
D19	0,251845
D114	0
D147	0
D161	0

Untuk contoh perhitungan kedua tabel diatas D8 diperoleh nilai 0,036533 yang didapatkan dari perhitungan pada rumus dibawah ini :

$$\text{Hasil Cosine} = \frac{\text{Jarak Q}}{\text{Panjang Vektor Q} * \text{Panjang vektor d8}}$$

$$\text{Hasil Cosine} = \frac{0,006598}{0,081228 * 2,223421}$$

$$\text{Hasil Cosine} = 0,036533$$

Selanjutnya dilakukan perankingan berdasarkan hasil *cosine* yang diperoleh. Setelah itu dilakukan pemilihan dokumen yang relevan dan tidak relevan sesuai dengan *query* yang sudah digunakan sebelumnya seperti bisa dilihat pada Tabel 4.11 di bawah ini

Tabel 4.10 Perankingan dokumen

Klasifikasi Q			RELEVAN
no	Jarak	ID	
1	0,251845063	D19	Iya
2	0,036532679	D8	iya
3	0	D114	Tidak
4	0	D147	Tidak
5	0	D161	Tidak

Hasil dari tabel 4.11 diatas selanjutnya dihitung berdasarkan tabel 2.6 *confusion matrix* seperti yang sudah dibahas pada bab II seperti bisa dilihat pada Tabel 4.12 di bawah ini

Tabel 4.11 Hasil *confusion matrix*

TP	2
TN	3
FP	0
FN	0

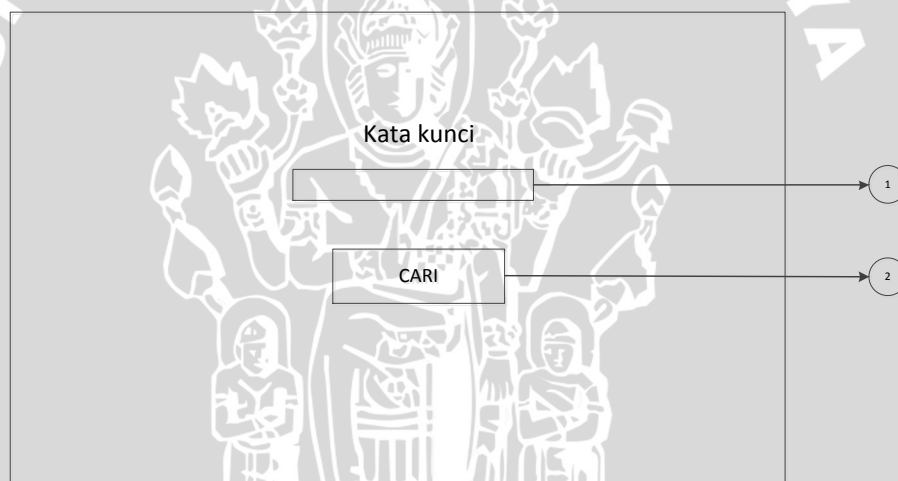
Selanjutnya hasil dari Tabel 4.12 di atas dihitung akurasi dengan *precision*, *recall*, *f-measure* dengan perincian seperti pada Tabel 4.13 di bawah ini

Tabel 4.12 Akurasi

Precision : $TP/(TP+FP)$	1
Recall : $TP/(TP+FN)$	1
F Measure (F1) : $2 * precision * recall / (precision + recall)$	1

4.4 Desain Antar Muka

Desain antar muka untuk sistem ini digunakan sebagai gambaran umum untuk mempermudah *user* guna memahami apa saja fitur utama yang ada pada sistem ini serta untuk apa saja fitur tersebut digunakan dalam sistem. Halaman ini akan muncul pertama kali saat sistem dijalankan. Untuk desain antar muka halaman beranda sistem bisa dilihat pada Gambar 4.10 berikut ini



Gambar 4.10 Desain Antar Muka Beranda

Keterangan Gambar 4.10:

1. *Field query* atau kata kunci digunakan untuk memasukkan kata kunci yang ingin dicari
2. Tombol Cari digunakan setelah *user* memasukkan *query* atau kata kunci.

Untuk selanjutnya perancangan desain antar muka untuk hasil pada halaman selanjutnya bisa dilihat pada Gambar 4.11 Desain Antar Muka Data berikut ini

Gambar 4.11 Desain Antar Muka Data

Fitur utama pada desain ini adalah fitur data yang digunakan untuk menampilkan keseluruhan data pada tahun 2014 sebanyak 207 data.

Selanjutnya perancangan desain antar muka untuk hasil pada halaman selanjutnya bisa dilihat pada Gambar 4.12 Desain Antar Muka Halaman Ketiga berikut

Gambar 4.12 Desain Antar Muka Tokenizing

Fitur utama pada desain antarmuka ini adalah *Tokenizing*. Fitur ini menampilkan data yang sudah terpisahkan menjadi kata-kata.

Desain antar muka untuk hasil pada halaman selanjutnya bisa dilihat pada Gambar 4.13 Desain Antar Muka *Filtering* berikut

Query	
Data	Tokenizing
Filtering	Stemming
TF.IDF.ICSDf	Cosine Similarity
Akurasi	
Show entries	Search
ID	Hasil Filtering
Previous 1 2 3 4 5 . . . Next	

Gambar 4.12 Desain Antar *Filtering*

Fitur utama pada desain antarmuka ini adalah *Filtering* yang berguna sebagai menghilangkan stoplist yang sudah ada pada database untuk data hasil *Tokenizing*.

Desain antar muka untuk hasil pada halaman selanjutnya bisa dilihat pada Gambar 4.14 Desain Antar Muka *Stemming* berikut

Query	
Data	Tokenizing
Filtering	Stemming
TF.IDF.ICSDf	Cosine Similarity
Akurasi	
Show entries	Search
ID	Hasil Stemming
Previous 1 2 3 4 5 . . . Next	

Gambar 4.13 Desain Antar Muka *Stemming*

Fitur utama pada desain antarmuka ini yaitu *Stemming*. Fitur ini digunakan untuk menghilangkan kata imbuhan ataupun akhiran dari hasil *Filtering* sehingga kata tersebut ditemukan kata dasarnya.

Desain antar muka untuk hasil pada halaman selanjutnya bisa dilihat pada Gambar 4.14 Desain Antar Muka TF.IDF.ICSDf berikut

Query

Data
Tokenizing
Filtering
Stemming
TF.IDF.ICSdF
Cosine Similarity
Akurasi

Show entries
Hasil TF.IDF.ICSdF
Search

No	Term	df	tf	D001	Dn	df	T	NT	T	NT	Csf	idf	ICSdF

Previous 1 2 3 4 5 ... Next

Gambar 4.14 Desain Antar Muka TF.IDF.ICSdF

Fitur utama pada halaman ini adalah TF.IDF.ICSdF yang digunakan sebagai metode pembobotan kata.

Desain antar muka untuk hasil pada halaman selanjutnya bisa dilihat pada Gambar 4.15 Desain Antar Muka *Cosine Similarity* berikut

Query

Data
Tokenizing
Filtering
Stemming
TF.IDF.ICSdF
Cosine Similarity
Akurasi

Show entries
Hasil Cosine Similarity
Search

ID	Hasil Cosine Similarity	Kelas	Hasil Cosine Sim

Previous 1 2 3 4 5 ... Next

Gambar 4.15 Desain Antar Muka *Cosine Similarity*

Fitur utama pada desain antarmuka ini adalah *cosine similarity* yang berguna untuk mencari kemiripan suatu dokumen serta merangking dokumen sesuai *query* yang dimasukkan.

Desain antar muka untuk hasil pada halaman selanjutnya bisa dilihat pada Gambar 4.16 Desain Antar Muka Akurasi Relevan berikut

Word Keys

Data
Tokenizing
Filtering
Stemming
TF.IDF.ICSDf
Cosine Similarity
Akurasi

Show entries
Search

No	Dokumen	Hasil Akurasi	Hasil Cosine	Relevan	Tidak Relevan
				☐ Relevan	☐ Tidak Relevan
				☐ Relevan	☐ Tidak Relevan
				☐ Relevan	☐ Tidak Relevan
				☐ Relevan	☐ Tidak Relevan
				☐ Relevan	☐ Tidak Relevan

Random
Submit

Previous 1 2 3 4 5 ... Next

Gambar 4.16 Desain Antar Muka Akurasi Relevan

Fitur utama pada desain antarmuka ini adalah akurasi yang berisi urutan teratas sampai terbawah dari dokumen menurut *query* yang dimasukan. *Radio botton* relevan dan tidak relevan digunakan untuk menentukan dokumen yang memang sesuai dan tidak sesuai dengan *query* awal . Selanjutnya hasil dari pemilihan tersebut digunakan menghitung *precission*, *recall*, *f-measure* dan akurasi selanjutnya. Tombol *random* digunakan untuk mengisi *radio botton* secara otomatis. Tombol *submit* digunakan untuk eksekusi ke halaman selanjutnya selanjutnya.

Desain antar muka untuk hasil pada halaman selanjutnya bisa dilihat pada Gambar 4.17 Desain Antar Muka Perhitungan Akurasi berikut

Query

Data
Tokenizing
Filtering
Stemming
TF.IDF.ICSDf
Cosine Similarity
Akurasi

Show entries
Search

Nama	Hasil Akurasi	Nilai

Gambar 4.17 Desain Antar Muka Perhitungan Akurasi

Fitur utama pada desain antarmuka ini adalah akurasi. Hasil untuk akurasi adalah perhitungan pada pengujian untuk menemukan dokumen relevan dan tidak relevan untuk selanjutnya dihitung *precision*, *recall* dan *f-measure*.

4.5 Perancangan Pengujian

Perancangan pengujian dilakukan untuk mengetahui bagaimana rancangan pengujian yang akan dilakukan pada sistem sudah sesuai apa belum dengan apa yang diharapkan. Perancangan pengujian untuk sistem ini menggunakan nilai *precision*, *recall* dan *f-measure*. Terdapat dua jenis pengujian yang akan dilakukan pada keseluruhan data pada tahun 2014 sebanyak 207 data. Perincian kedua perancangan pengujian dihitung dengan metode TF.IDF.ICF dan yang kedua menggunakan TF.IDF.ICSDf.

Perancangan pengujian pertama yaitu yang pertama dilakukan dengan menggunakan 3 *query* yang berbeda yang terdiri dari lebih dari 2 *term* atau kata dan akan dibandingkan dengan 3 *query* yang terdiri dari 3 *term* atau kata dengan catatan mengandung *term* pada pengujian pertama proses pertama. Masing-masing dihitung dengan menggunakan metode TF.IDF.ICF dan TF.IDF.ICSDf.

Tujuan pengujian dilakukan dengan *query* awal dan *query* akhir dimana *query* kedua masih mengandung *term* pada *query* pertama adalah untuk mengetahui hasil dari perubahan nilai *f-measure* akan tetap, berkurang atau bertambah. Pada penelitian ini diharapkan nilai *f-measure* pada pengujian akan semakin meningkat. Ini dikarenakan semakin bertambah nilai *f-measure* maka akan berpengaruh pada kenaikan tingkat perangkungan dokumen itu sendiri. Perancangan pengujian pertama ke n awal dan akhir bisa dilihat pada Tabel 4.13 dibawah

Tabel 4.13 Perancangan pengujian pertama ke n *query* awal dan akhir metode TF.IDF atau TF.IDF.ICSDf

Indikator	Query Awal	Query Akhir
Precision		
Recall		
F Measure		

Tabel diatas merupakan tabel hasil akhir untuk *precision*, *recall* dan F1 dari perhitungan metode TF.IDF.ICF maupun TF.IDF.ICSDf menggunakan *query* awal yang terdiri dari lebih dari 2 *term* atau kata dan *query* akhir yang terdiri dari lebih dari 3 kata yang mengandung *term* sebelumnya pada *query* awal.

Tahap selanjutnya dari pengujian diatas adalah mengukur tingkat *relevancy* dokumen dengan menggunakan hasil dari *cosine similarity* yang telah dihitung sebelumnya dengan *query* tertentu. Hasil cosine digunakan sebagai acuan adalah karena *cosine similarity* mengadopsi aturan dimana *query* dan dokumen direpresentasikan sebagai vektor. Jika semakin kecil sudut yang diperoleh dokumen terhadap *query* tertentu maka semakin mirip dokumen tersebut. Urutan perangkungan didasarkan pada hasil dari sudut terkecil yang mendekati *query*.

BAB 5 IMPLEMENTASI

Bab ini membahas tentang implementasi sistem temu kembali informasi dokumen ide kreatif berbahasa Indonesia menggunakan *vector space model* dengan metode *inverse class space density frequency*. Pembahasan untuk bab ini meliputi spesifikasi sistem, batasan implementasi sistem, implementasi algoritma dan implementasi *user interface*.

5.1 Spesifikasi Sistem

Hasil analisis dan perancangan yang telah dilakukan pada Bab IV menjadi acuan untuk melakukan implementasi pada bab ini. Spesifikasi sistem terdiri dari spesifikasi perangkat keras dan spesifikasi perangkat lunak.

5.1.1 Spesifikasi Perangkat Keras

Spesifikasi perangkat keras yang digunakan untuk pengembangan pada sistem ini dijelaskan pada tabel 5.1 di bawah

Tabel 5.1 Spesifikasi Perangkat Keras

Nama Komponen	Spesifikasi
Prosesor	AMD A10-5750M 2.5GHz
Memori	4 GB
Hardisk	1 TB

5.1.2 Spesifikasi Perangkat Lunak

Spesifikasi perangkat lunak yang digunakan untuk pengembangan pada sistem ini dijelaskan pada tabel 5.2 di bawah

Tabel 5.2 Spesifikasi Perangkat Lunak

Nama Komponen	Spesifikasi
Sistem Operasi	Windows 8.1
Bahasa Pemrograman	Javascript, Php, MySql
Tools Pemrograman	Sublime, Notepad++, Chrome

5.2 Batasan Implementasi Sistem

Batasan implementasi sistem ditampilkan agar penelitian ini memiliki ruang lingkup yang jelas dalam mengimplementasikan sistem. Terdapat beberapa batasan yang membatasi pengembangan sistem ini di antaranya

BAB 1 *Input* untuk sistem adalah *query* sebagai masukan *user* untuk data ide kreatif tahun 2014.

BAB 2 *Output* berupa perankingan dokumen sesuai query masukan *user*.

BAB 3 Pengembangan sistem dilakukan dengan menggunakan bahasa pemrograman PHP dan Javascript.

BAB 4 Algoritma yang digunakan adalah TF.IDF.ICSDf, *Cosine Similarity*

BAB 5 *Stemming* menggunakan algoritma Porter

5.3 Implementasi Algoritma

Implementasi Algoritma untuk sistem ini terdiri dari beberapa proses yaitu *Filtering*, *Stemming*, ICSDf dan *Cosine Similarity*.

5.3.1 Implementasi Filtering

Implementasi *Filtering* bisa dilihat pada *Sourcode* 5.1 berikut

```

1  for ($k = 0; $k < sizeof($token); $k++) {
2      echo '<tr><td>'. $dokumen[$k][2]. '</td>';
3      $print = "";
4      $count = 0;
5      for ($l = 0; $l < sizeof($token[$k]); $l++) {
6          $cek = true;
7          if (in_array($token[$k][$l], $wordlist)) {
8              $cek = false;
9          }
10         if ($cek == true) {
11             $print .= $token[$k][$l]. '; ';
12             $filtering[$k][$count] = $token[$k][$l];
13             $count++;
14         }
15     }
16     echo '<td>'. substr($print, 0, -2). '</td></tr>';
17 }

```

Source code 5.1 Implementasi Filtering

Untuk penjelasan baris pada *sourcecode* diatas dapat dilihat di bawah ini

Baris 1 = Dilakukan perulangan sebanyak token pada dokumen (token) berisi dokumen yang sudah dipecah

Baris2 = Mencetak id (D001,D002 dsb)

Baris 3-4 = inisialisasi

Baris 5 = Mencari perulangan sebanyak kolom yang berisi token dari baris 1

Baris 6 = Membuat inisialisasi *default = true*

Baris 7 = Mengecek apakah kata pada stoplist ada pada wordlist atau tidak

Baris 8 = Inisialisasi jika ada *stoplist* maka *false*

Baris 10-15 = Mengecek jika *true* maka disimpan pada \$print dan ditambah “;” lalu disimpan juga di variable \$filtering[\$k][\$count]

Baris 16 = Mencetak yang disimpan pada \$print dan menghilangkan 2 karakter string pada kata terakhir

5.3.2 Implementasi Stemming

Implementasi *Stemming* bisa dilihat pada *Source code* 5.2 berikut

```

1  if (in_array(strtolower($stemming[$j][$k]), $kata) ||
2  strtoupper($stemming[$j][$k]) == $stemming[$j][$k]) {
3      $print .= $stemming[$j][$k].'; ';
4      continue;
5  }
6  if (in_array(strtolower($stemming[$j][$k]), $kata) ||
7  strtoupper($stemming[$j][$k]) == $stemming[$j][$k]) {
8      $print .= $stemming[$j][$k].'; ';
9      continue; // utk mencari kata dasar }
10 if(substr(strtolower($stemming[$j][$k]),
11 strlen(strtolower($stemming[$j][$k]))-3,
12 strlen(strtolower($stemming[$j][$k]))) == 'kah') {
13 $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 0
14 , (strlen(strtolower($stemming[$j][$k]))-3));
15 elseif(substr(strtolower($stemming[$j][$k]),
16 (strlen(strtolower($stemming[$j][$k]))-3,
17 strlen(strtolower($stemming[$j][$k]))) == 'lah') {
18 $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 0
19 , (strlen(strtolower($stemming[$j][$k]))-3));
20 elseif(substr(strtolower($stemming[$j][$k]),
21 (strlen(strtolower($stemming[$j][$k]))-3,
22 strlen(strtolower($stemming[$j][$k]))) == 'pun') {
23 $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 0
24 , (strlen(strtolower($stemming[$j][$k]))-3));
25 }
26 //Inflectional Possesive Pronoun
27 if(substr(strtolower($stemming[$j][$k]),
28 (strlen(strtolower($stemming[$j][$k]))-2,
29 strlen(strtolower($stemming[$j][$k]))) == 'ku') {
30 $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 0
31 , (strlen(strtolower($stemming[$j][$k]))-2));}
32 elseif(substr(strtolower($stemming[$j][$k]),
33 (strlen(strtolower($stemming[$j][$k]))-2,
34 strlen(strtolower($stemming[$j][$k]))) == 'mu') {
35 $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 0
36 , (strlen(strtolower($stemming[$j][$k]))-2));}
37 elseif(substr(strtolower($stemming[$j][$k]),
38 (strlen(strtolower($stemming[$j][$k]))-3,
39 strlen(strtolower($stemming[$j][$k]))) == 'nya') {
40 $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 0
41 , (strlen(strtolower($stemming[$j][$k]))-3));
42 }
43 // First Order Derivational Prefix
44 if (substr(strtolower($stemming[$j][$k]), 0, 4) == 'meng')
45 {
46 $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]),4,s
47 trlen(strtolower($stemming[$j][$k])));$DiMengTer = true;}
48 elseif (substr(strtolower($stemming[$j][$k]), 0,
49 4) == 'meny') {
50 $stemming[$j][$k]='s'.substr(strtolower($stemming[$j][$k])
51 , 4 , strlen(strtolower($stemming[$j][$k])));}
52 elseif (substr(strtolower($stemming[$j][$k]), 0,

```

```

53     3) == 'mem')
54     {
55     if (substr(strtolower($stemming[$j][$k]), 3, 1) == 'a' ||
56     substr(strtolower($stemming[$j][$k]), 3, 1) == 'e' ||
57     substr(strtolower($stemming[$j][$k]), 3, 1) == 'i'
58     || substr(strtolower($stemming[$j][$k]), 3, 1) == 'o' ||
59     substr(strtolower($stemming[$j][$k]), 3, 1) == 'u') {
60     $stemming[$j][$k]='p'.substr(strtolower($stemming[$j][$k])
61     , 3 , strlen(strtolower($stemming[$j][$k])));}
62     else{$stemming[$j][$k]=
63     substr(strtolower($stemming[$j][$k]),3 ,
63     strlen (strtolower ($stemming[$j][$k])));}
64     elseif (substr(strtolower($stemming[$j][$k]), 0, 3) ==
65     'men') {
66     if (substr(strtolower($stemming[$j][$k]), 3, 1) == 'a' ||
67     substr(strtolower($stemming[$j][$k]), 3, 1) == 'e' ||
68     substr(strtolower($stemming[$j][$k]), 3, 1) == 'i'
69     || substr(strtolower($stemming[$j][$k]), 3, 1) == 'o' ||
70     substr(strtolower($stemming[$j][$k]), 3, 1) == 'u') {
71     $stemming[$j][$k]='t'.substr(strtolower($stemming[$j][$k])
72     , 3 , strlen(strtolower($stemming[$j][$k])));}
73     else{$stemming[$j][$k]=
74     substr(strtolower($stemming[$j][$k]),3 ,
75     strlen(strtolower($stemming[$j][$k])));}
76     elseif (substr(strtolower($stemming[$j][$k]), 0, 2) ==
77     'me'){$stemming[$j][$k]=substr(strtolower($stemming[$j][$k]
78     ), 2 , strlen(strtolower($stemming[$j][$k])));}
79     elseif (substr(strtolower($stemming[$j][$k]), 0, 4) ==
80     'peng') {
81     $stemming[$j][$k]= substr(strtolower($stemming[$j][$k]), 4
82     , strlen(strtolower($stemming[$j][$k])));}
83     $KePeng = true;
84     $BerKePeng = true;}
85     elseif (substr(strtolower($stemming[$j][$k]), 0, 4) ==
86     'peny') {
87     $stemming[$j][$k]='s'.substr(strtolower($stemming[$j][$k])
88     , 4 , strlen(strtolower($stemming[$j][$k])));}
89     elseif(substr(strtolower($stemming[$j][$k]), 0, 3) ==
90     'pen') {
91     if (substr(strtolower($stemming[$j][$k]), 3, 1) ==
92     'd' && substr(strtolower($stemming[$j][$k]),
93     (strlen(strtolower($stemming[$j][$k]))-3),
94     strlen(strtolower($stemming[$j][$k])))) == 'kan') {
95     $stemming[$j][$k]='t'.substr (strtolower
96     ($stemming[$j][$k]), 0 ,
97     (strlen(strtolower($stemming[$j][$k]))-2));}
98     $stemming[$j][$k]=
99     substr(strtolower($stemming[$j][$k]), 3 ,
100     strlen(strtolower($stemming[$j][$k])));}
101     elseif (substr(strtolower($stemming[$j][$k]), 0, 3) ==
102     'pem')
103     {
104     if (substr(strtolower($stemming[$j][$k]),3, 1 == 'a') ||
105     substr(strtolower($stemming[$j][$k]), 3, 1 == 'e')
106     ||substr(strtolower($stemming[$j][$k]),3,1=='i')||substr(s
107     trtolower($stemming[$j][$k]),3,1=='o')||substr(strtolower(
108     $stemming[$j][$k]), 3, 1 == 'u')) {
109     $stemming[$j][$k]='p'.substr(strtolower($stemming[$j][$k])
110     , 3 , strlen(strtolower($stemming[$j][$k])));}

```

```

111     else {
112         $stemming[$j][$k]=
113         substr(strtolower($stemming[$j][$k]), 3 ,
114         strlen(strtolower($stemming[$j][$k])));}
115     elseif (substr(strtolower($stemming[$j][$k]), 0, 2)
116     =='di'){
117         $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 2
118         , strlen(strtolower($stemming[$j][$k])));
119         $DiMengTer = true;}
120     elseif (substr(strtolower($stemming[$j][$k]), 0, 3) ==
121     'ter') {
122         $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 3
123         , strlen(strtolower($stemming[$j][$k])));
124         $DiMengTer = true;}
125     elseif (substr(strtolower($stemming[$j][$k]), 0, 2) ==
126     'ke') {
127         $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 2
128         , strlen(strtolower($stemming[$j][$k])));
129         $KePeng = true;
130         $BerKePeng = true;
131     }
132     else{
133         $sempataA = true;} if ($sempataA == true) {
134     //Second Order Derivational Prefix
135     if (substr(strtolower($stemming[$j][$k]), 0, 3) == 'ber') {
136         $stemming[$j][$k]= substr(strtolower($stemming[$j][$k]), 3
137         , strlen(strtolower($stemming[$j][$k])));
138         $BerKePeng = true;}
139         elseif (substr(strtolower($stemming[$j][$k]), 0,
140         3) == 'bel') {
141             $stemming[$j][$k]=
142             substr(strtolower($stemming[$j][$k]), 3 ,
143             strlen(strtolower($stemming[$j][$k])));}
144         elseif (substr(strtolower($stemming[$j][$k]), 0,
145         2) == 'be') {
146             $stemming[$j][$k]=
147             substr(strtolower($stemming[$j][$k]), 2 ,
148             strlen(strtolower($stemming[$j][$k])));}
149         elseif (substr(strtolower($stemming[$j][$k]), 0,
150         3) == 'per') {
151             $stemming[$j][$k]= substr(strtolower
152             ($stemming[$j][$k]),3,
153             strlen(strtolower($stemming[$j][$k])));}
154         elseif (substr(strtolower($stemming[$j][$k]), 0,
155         3) == 'pel') {
156             $stemming[$j][$k]=
157             substr(strtolower($stemming[$j][$k]), 3 ,
158             strlen(strtolower($stemming[$j][$k])));}
159         elseif (substr(strtolower($stemming[$j][$k]), 0,
160         2) == 'pe') {
161             $stemming[$j][$k]=
162             substr(strtolower($stemming[$j][$k]), 2 ,
163             strlen(strtolower($stemming[$j][$k])));}
164     //Derrivational Suffix
165     if (substr(strtolower($stemming[$j][$k]), (strlen(strtolower
166     ($stemming[$j][$k]))-3),
167     strlen(strtolower($stemming[$j][$k])))) == 'kan' &&
168     !$KePeng) {
169

```

```

170 $stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 0
171 , (strlen(strtolower($stemming[$j][$k]))-3));}
172 elseif(substr(strtolower($stemming[$j][$k]),
173 strlen(strtolower($stemming[$j][$k]))-2),
174 strlen(strtolower($stemming[$j][$k]))) == 'an' &&
175 !$DiMengTer) {
1751 $stemming[$j][$k]=
177 substr(strtolower($stemming[$j][$k]), 0 ,
178 (strlen(strtolower($stemming[$j][$k]))-2));
179 elseif(substr(strtolower($stemming[$j][$k]),
180 (strlen(strtolower($stemming[$j][$k]))-1),
181 strlen(strtolower($stemming[$j][$k]))) == 'i' &&
182 !$BerKePeng) {
183 $stemming[$j][$k]=
184 substr(strtolower($stemming[$j][$k]), 0 ,
185 (strlen(strtolower($stemming[$j][$k]))-1));}
186 else {
187 //Derivational Suffix
188 if(substr(strtolower($stemming[$j][$k]), (strlen(strtolower
189 ($stemming[$j][$k]))-3),
190 strlen(strtolower($stemming[$j][$k]))) == 'kan' &&
191 !$KePeng) {
192 $stemming[$j][$k]= substr(strtolower($stemming[$j][$k]), 0
193 , (strlen(strtolower($stemming[$j][$k]))-3));}
194 elseif(substr(strtolower($stemming[$j][$k]),
195 (strlen(strtolower($stemming[$j][$k]))-2),
196 strlen(strtolower($stemming[$j][$k]))) == 'an'
197 && !$DiMengTer) {
198 $stemming[$j][$k]=
199 substr(strtolower($stemming[$j][$k]), 0 ,
200 (strlen(strtolower($stemming[$j][$k]))-2));}
201 elseif(substr(strtolower($stemming[$j][$k]),
202 (strlen(strtolower($stemming[$j][$k]))-1),
203 strlen(strtolower($stemming[$j][$k]))) == 'i' &&
204 !$BerKePeng) {$stemming[$j][$k]=
205 substr(strtolower($stemming[$j][$k]), 0 ,
206 (strlen(strtolower($stemming[$j][$k]))-1));}
207 else {$print .= $stemming[$j][$k].'; ';
208 continue;}
209 //Second Order Derivational Prefix 5b
210 if (substr(strtolower($stemming[$j][$k]), 0, 3) == 'ber')
211 {$stemming[$j][$k]=substr(strtolower($stemming[$j][$k]), 3
212 , strlen(strtolower($stemming[$j][$k])));}
213 elseif (substr(strtolower($stemming[$j][$k]),
214 0, 3) == 'bel') {
215 $stemming[$j][$k]=
216 substr(strtolower($stemming[$j][$k]), 3 ,
217 strlen(strtolower($stemming[$j][$k])));}
218 elseif (substr(strtolower($stemming[$j][$k]),
219 0, 2) == 'be') {
220 $stemming[$j][$k]=
221 substr(strtolower($stemming[$j][$k]), 2 ,
222 strlen(strtolower($stemming[$j][$k])));}
223 elseif (substr(strtolower($stemming[$j][$k]),
224 0, 3) == 'per') {
225 $stemming[$j][$k]=
226 substr(strtolower($stemming[$j][$k]), 3 ,
227 strlen(strtolower($stemming[$j][$k])));}
228 elseif (substr(strtolower($stemming[$j][$k]),

```

229	0, 3) == 'pel') {
230	\$stemming[\$j][\$k]=
231	substr(strtolower(\$stemming[\$j][\$k]), 3 ,
232	strlen(strtolower(\$stemming[\$j][\$k])));}
233	elseif (substr(strtolower(\$stemming[\$j][\$k]),
234	0, 2) == 'pe') {
235	\$stemming[\$j][\$k] =
236	substr(strtolower(\$stemming[\$j][\$k]), 2 ,
237	strlen(strtolower(\$stemming[\$j][\$k])));
	}

Source code 5.2 Implementasi Stemming

Untuk penjelasan baris pada *source code* diatas dapat dilihat di bawah

- Baris 1-5 = Mencari kata dasar yang ada lalu disimpan pada \$sprint lalu melanjutkan pada tahap berikutnya
- Baris 6-25 = Menghapus akhiran (kah, lah, pun) pada tiap kata
- Baris 26-42 = Menghilangkan akhiran (ku, mu, nya)
- Baris 43-133 = Menghapus awalan pertama(meng, meny, men, mem, me, peng, peny, pen, pem, di, ter, ke) dan misalkan tidak ada imbuhan didepan \$smpataA maka *true*
- Baris 134-164 = Menghapus awalan kedua (ber, be, per, pel, pe)
- Baris 166-186 = Menghapus akhiran depan (kan) saat bukan anggota (ke, peng), Menghapus akhiran (an) saat bukan anggota (di, meng, ter, Menghapus akhiran (i) saat bukan anggota (ber, ke, peng)
- Baris 188-208 = Menghapus akhiran depan (kan) saat bukan anggota (ke, peng), Menghapus akhiran (an) saat bukan anggota (di, meng , ter ,Menghapus akhiran (i) saat bukan anggota (ber, ke , peng) dengan catatan saat awalan tidak dihapus maka dianggap kata dasar
- Baris 210-237 = Menghapus awalan (ber, bel, be, per, pel, pe)

5.3.3 Implementasi ICSdf

Implementasi ICSdf bisa dilihat pada *Source code* 5.3 berikut

1	for (\$k = 0; \$k < \$count; \$k++) {
2	if (empty(\$pembobotan[\$k])) {
3	continue;
4	}
5	\$index3++;
6	\$index = 0;
7	\$tfidf[\$k][\$index] = \$pembobotan[\$k];
8	\$index += 1;
9	\$ccek = 0;
10	for (\$q = 0; \$q < sizeof(\$cari); \$q++) {
11	if (\$cari[\$q] ==
12	strtolower(\$pembobotan[\$k])) {
13	\$ccek += 1;
14	}
15	}

```

16         if ($cek != 0) {
17             echo '<tr><td style="background-
18 color:green">'. ($index3) .'</td><td
19 style="background-color:green">'.
20 $pembobotan[$k] .'</td>';
21             echo '<td style="background-
22 color:green">'. $cek .'</td>';
23         } else {
24             echo '<tr><td>'. ($index3) .'</td><td>'.
25 $pembobotan[$k] .'</td>';
26             echo '<td>'. $cek .'</td>';}
27             $tfidf[$k][$index] = $cek;
28             $index += 1;
29             $df = 0;
30             $wck = array();
31             for ($w = 0; $w < 2; $w++) {
32                 $wck[$w] = 0; }
33             for ($r = 0; $r < sizeof($stemming); $r++) {
34                 $angka[$k][$r] = 0;
35                 $log = 0;
36                 for ($s = 0; $s < sizeof($stemming[$r]);
37 $s++) {
38                     if (strtolower($stemming[$r][$s]) ==
39 strtolower($pembobotan[$k])) {
40                         $angka[$k][$r] += 1;}}
41                 if ($angka[$k][$r] > 0) {
42                     $df += 1;}}
43                 if ($cek != 0) {
44                     echo '<td style="background
45 color:green">'. $angka[$k][$r] .'</td>';
46                 } else {
47                     echo '<td>'. $angka[$k][$r] .'</td>';}
48                 if ($dokumen[$r][0] == "T" &&
49 $angka[$k][$r] > 0) {
50                     $wck[0] = 1;}}
51                 if ($dokumen[$r][0] == "NT" &&
52 $angka[$k][$r] > 0) {
53                     $wck[1] = 1;}}
54                 $log = ($df == 0) ? 0 :
55 log10(sizeof($stemming) / $df);
56                 $tfidf[$k][$index] = $df;
57                 $index += 1;
58                 $wckT = ($wck[0] / $jum['T']);
59                 $wckNT = ($wck[1] / $jum['NT']);
60                 $csdf = $wckT + $wckNT;
61                 $ICSdF = log10(2/$csdf);
62                 if ($cek != 0) {
63                     echo '<td style="background-
64 color:green">'. $df .'</td>';
65                     echo '<td style="background-
66 color:green">'. $log .'</td>';
67                     echo '<td style="background-
68 color:green">'. $wck[0] .'</td>';
69                     echo '<td style="background-
70 color:green">'. $wck[1] .'</td>';
71                     echo '<td style="background-
72 color:green">'. $wckT .'</td>';
73                     echo '<td style="background-
74 color:green">'. $wckNT .'</td>';
75

```

```

76         echo '<td style="background-
77         color:green">'. $csdf .'

```

Source code 5.3 Implementasi ICSdF

Untuk penjelasan baris pada *source code* diatas dapat dilihat di bawah

- Baris 1 = Perulangan sebanyak jumlah kata
- Baris 2-4 = Melewati kata yang sudah pernah dihitung sebelumnya
- Baris 5-9 = Inisialisasi
- Baris 10-15 = Mencari berapa banyak *query*
- Baris 16-26 = Mencetak dokumen yang masuk ranking dengan warna hijau dan untuk baris berikutnya mencetak warna *default* untuk dokumen yang tidak masuk ranking
- Baris 27-32 = inisialisasi
- Baris 33 = Perulangan sejumlah banyak *stemming*

Baris 34-35	= Inisialisasi
Baris 36-40	= Mencari <i>Term Frequency</i>
Baris 41-42	= Mencari <i>Document Frequency</i>
Baris 43-47	= Mencetak warna hijau hasil term frequency saat termasuk sebagai kata yang dicari dan untuk baris berikutnya mencetak warna default untuk dokumen yang tidak masuk ranking
Baris 48-53	= Jika kata tersebut ada di Teknik dan < 0 maka Wck Teknik 1 dan Jika kata tersebut ada di NonTeknik dan < 0 maka Wck NonTeknik 1
Baris 54-62	= Inisialisasi idf, icpdf, csdf
Baris 63-88	= Mencetak hasil inisialisasi
Baris 89-90	= Inisialisasi
Baris 91-99	= Mencetak warna hijau hasil <i>weight</i> saat termasuk sebagai kata yang dicari dan untuk baris berikutnya mencetak warna <i>default</i> untuk dokumen yang tidak masuk rangking
Baris 101-103	= Inisialisasi
Baris 104-116	= Mencetak hasil <i>weight</i>

5.3.4 Implementasi *Cosine Similarity*

Implementasi *Cosine Similarity* 1 bisa dilihat pada *Source code* 5.4 berikut

```

1  for ($i = 0; $i <= $hitung; $i++) {
2      if ($i != $hitung) {
3          $jarak[$i] = 0;
4          $panjang[$i] = 0;
5          for ($j = 0; $j < $count; $j++) {
6              if (empty($tfidf[$j])) {
7                  continue;
8              }
9              if ($i != $hitung) {
10                 $panjang[$i] += pow($tfidf[$j][4 + $i], 2);
11                 if ($tfidf[$j][4] != 0 && $tfidf[$j][5 + $i] != 0)
12                 {
13                     $jarak[$i] += ($tfidf[$j][4] * $tfidf[$j][5 +
14                     $i]);
15                     // echo $tfidf[$j][4]. " * ". $tfidf[$j][5 +
16                     $i]. "<br>";
17                 }
18             }
19         }
20     }
21 }
```

Source code 5.4 Implementasi *Cosine Similarity* 1

Untuk penjelasan baris pada *source code* diatas dapat dilihat di bawah

Baris 1	= Perulangan sebanyak dokumen
Baris 2-4	= Inisialisasi jarak dan panjang
Baris 5-15	= Menghitung jarak dan panjang

5.4 Desain Antarmuka

Desain Antarmuka berisi halaman-halaman keseluruhan dari sistem yang berisi fitur utama pada setiap halamannya. Fitur tersebut antara lain fitur data, fitur *tokenizing*, fitur *filtering*, fitur *stemming*, fitur *cosine similarity* dan fitur akurasi.

5.4.1 Antarmuka Beranda

Antarmuka beranda bisa dilihat pada Gambar 5.1 di bawah



Gambar 5.1 Antarmuka beranda

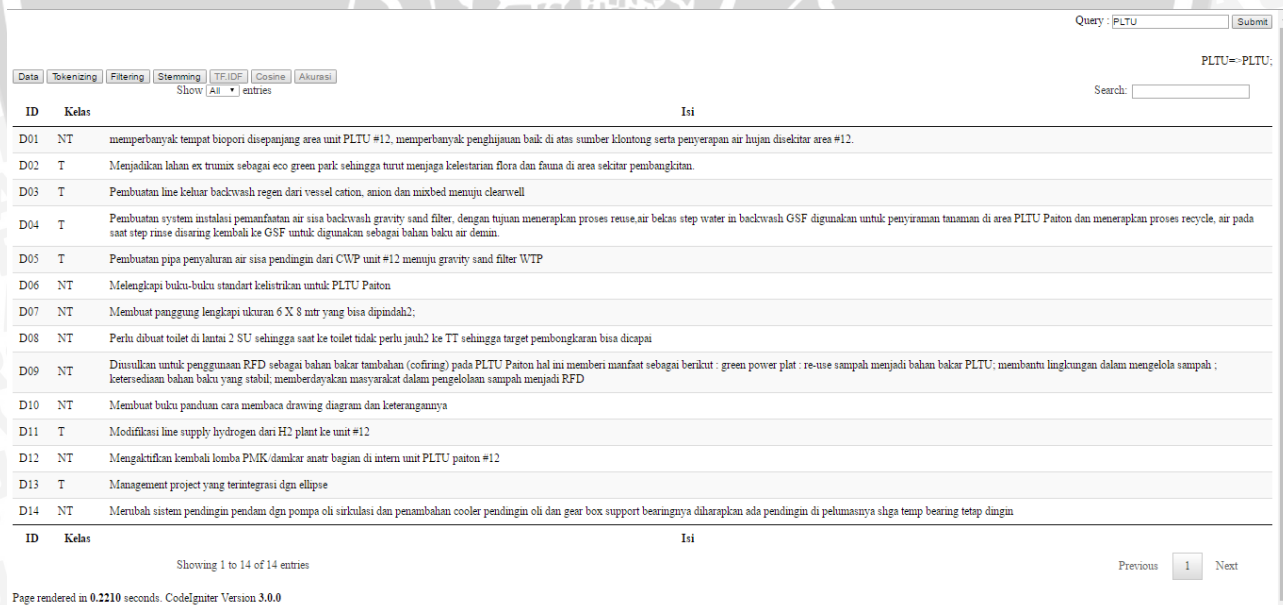
Keterangan Gambar 5.1 bisa dilihat pada Tabel 5.3 di bawah

Tabel 5.3 Keterangan Antarmuka Beranda

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya

5.4.2 Antarmuka Data

Antarmuka Data bisa dilihat pada Gambar 5.2 di bawah



ID	Kelas	Isi
D01	NT	memperbanyak tempat biopori disepanjang area unit PLTU #12, memperbanyak penghijauan bak di atas sumber idoutong serta penyerapan air hujan disekitar area #12.
D02	T	Menjadikan lahan ex trumix sebagai eco green park sehingga turut menjaga kelestarian flora dan fauna di area sekitar pembangkitan.
D03	T	Pembuatan line kehar backwash regen dari vessel cotton, amon dan mixed menuju clearwell
D04	T	Pembuatan system instalasi penumfatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paton dan menerapkan proses recycle, air pada saat step rinse disaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.
D05	T	Pembuatan pipa penyaluran air sisa pendingin dari CWP unit #12 menuju gravity sand filter WTP
D06	NT	Melengkapi buku-buku standart kelistrikan untuk PLTU Paton
D07	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;
D08	NT	Perlu dibuat toilet di lantai 2 SU sehingga saat ke toilet tidak perlu jauh2 ke TT sehingga target pembongkaran bisa dicapai
D09	NT	Ditentukan untuk penggunaan RFD sebagai bahan bakar tambahan (cofiring) pada PLTU Paton hal ini memberi manfaat sebagai berikut : green power plat : re-use sampah menjadi bahan bakar PLTU; membantu lingkungan dalam mengelola sampah ; ketersediaan bahan baku yang stabil; memberdayakan masyarakat dalam pengelolaan sampah menjadi RFD
D10	NT	Membuat buku panduan cara membaca drawing diagram dan keterangannya
D11	T	Modifikasi line supply hydrogen dari H2 plant ke unit #12
D12	NT	Mengaktifkan kembali lomba PMK/damkar anstr bagian di intern unit PLTU paton #12
D13	T	Management project yang terintegrasi dgn ellipse
D14	NT	Merubah sistem pendingin pendam dgn pompa oli sirkulasi dan penambahan cooler pendingin oli dan gear box support bearingnya diharapakan ada pendingin di perumahan shga temp bearing tetap dingin

Gambar 5.2 Antarmuka Data

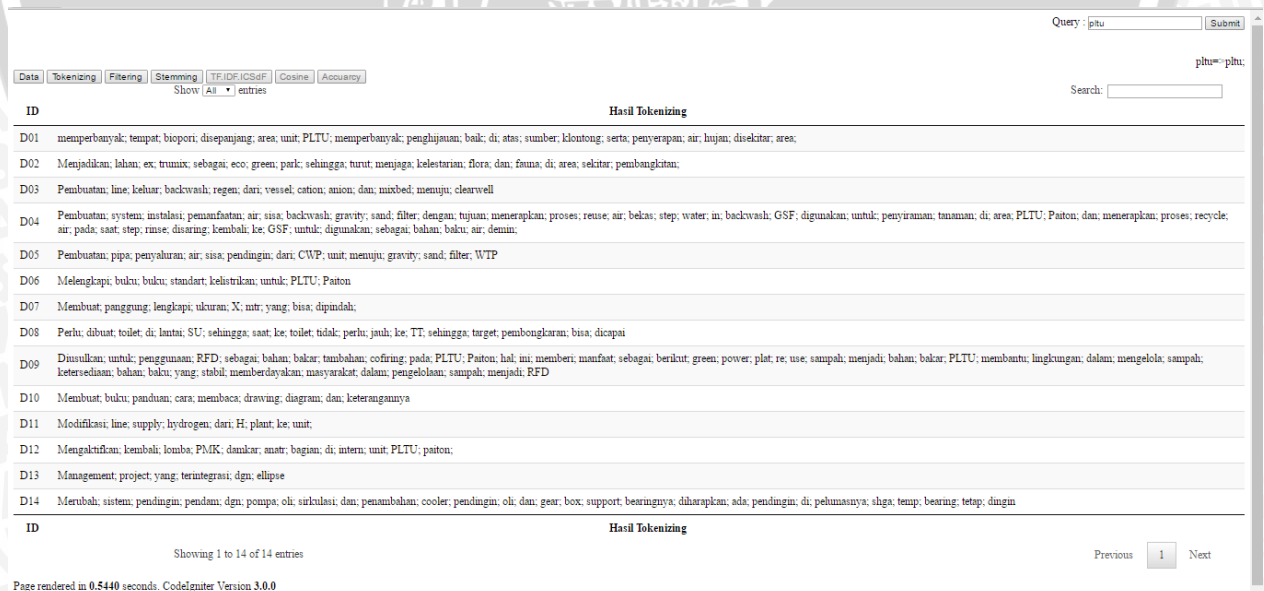
Keterangan Gambar 5.2 bisa dilihat pada Tabel 5.4 di bawah

Tabel 5.4 Antarmuka Data

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen.
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya.
<i>Data</i>	Sebagai tempat keseluruhan data tahun 2014.
<i>Tokenizing</i>	Sebagai tempat data yang sudah berupa kata-kata.
<i>Filtering</i>	Sebagai tempat data hasil <i>tokenizing</i> yang sudah dihilangkan <i>stoplist</i> nya.
<i>Stemming</i>	Sebagai tempat data hasil <i>filtering</i> yang sudah dihilangkan kata imbuhan dan akhirnya
TF.IDF.ICSDf	Sebagai tempat kata hasil <i>stemming</i> yang sudah dihitung bobotnya.
<i>Cosine</i>	Sebagai tempat data hasil TF.IDF.ICSDf untuk dihitung kemiripannya dengan <i>query inputan</i> pertama yang selanjutnya digunakan untuk meranking dokumen.
<i>Accurasy</i>	Sebagai tempat data yang sudah dihitung akurasi dengan <i>precision</i> , <i>recall</i> , <i>f-measure</i> dan akurasi

5.4.3 Antarmuka *Tokenizing*

Antarmuka *Tokenizing* bisa dilihat pada Gambar 5.3 di bawah



Gambar 5.3 Antarmuka *Tokenizing*

Keterangan Gambar 5.3 bisa dilihat pada Tabel 5.5 di bawah

Tabel 5.5 Antarmuka *Tokenizing*

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen.
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya.
<i>Data</i>	Sebagai tempat keseluruhan data tahun 2014.
<i>Tokenizing</i>	Sebagai tempat data yang sudah berupa kata-kata.
<i>Filtering</i>	Sebagai tempat data hasil <i>tokenizing</i> yang sudah dihilangkan <i>stoplist</i> nya.
<i>Stemming</i>	Sebagai tempat data hasil <i>filtering</i> yang sudah dihilangkan kata imbuhan dan akhirnya
TF.IDF.ICSDf	Sebagai tempat kata hasil <i>stemming</i> yang sudah dihitung bobotnya.
<i>Cosine</i>	Sebagai tempat data hasil TF.IDF.ICSDf untuk dihitung kemiripannya dengan <i>query inputan</i> pertama yang selanjutnya digunakan untuk meranking dokumen.
<i>Accurasy</i>	Sebagai tempat data yang sudah dihitung akurasi dengan <i>precision</i> , <i>recall</i> , <i>f-measure</i> dan akurasi

5.4.4 Antarmuka *Filtering*

Antarmuka *Filtering* bisa dilihat pada Gambar 5.4 di bawah

Gambar 5.4 Antarmuka *Filtering*

Keterangan Gambar 5.4 bisa dilihat pada Tabel 5.6 di bawah

Tabel 5.6 Antarmuka Filtering

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen.
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya.
<i>Data</i>	Sebagai tempat keseluruhan data tahun 2014.
<i>Tokenizing</i>	Sebagai tempat data yang sudah berupa kata-kata.
<i>Filtering</i>	Sebagai tempat data hasil <i>tokenizing</i> yang sudah dihilangkan <i>stoplist</i> nya.
<i>Stemming</i>	Sebagai tempat data hasil <i>filtering</i> yang sudah dihilangkan kata imbuhan dan akhirnya
TF.IDF.ICSDf	Sebagai tempat kata hasil <i>stemming</i> yang sudah dihitung bobotnya.
<i>Cosine</i>	Sebagai tempat data hasil TF.IDF.ICSDf untuk dihitung kemiripannya dengan <i>query inputan</i> pertama yang selanjutnya digunakan untuk meranking dokumen.
<i>Accurasy</i>	Sebagai tempat data yang sudah dihitung akurasi dengan <i>precision</i> , <i>recall</i> , <i>f-measure</i> dan akurasi

5.4.5 Antarmuka Stemming

Antarmuka *Stemming* bisa dilihat pada Gambar 5.5 di bawah

The screenshot shows the 'Antarmuka Stemming' interface. At the top, there are tabs for 'Data', 'Tokenizing', 'Filtering', 'Stemming', 'TF.IDF.ICSDf', 'Cosine', and 'Accuracy'. The 'Stemming' tab is selected. Below the tabs, there is a search bar with the text 'Query: jitu' and a 'Submit' button. To the right of the search bar, there is a 'Show All' button and a 'Entries' label. Below the search bar, there is a table with the following data:

ID	Hasil Stemming
D01	perbanyak; biopor; sepaung; are; usat; PLTU; perbanyak; biopor; sumber; kloning; sepa; air; biopor; ekster; are;
D02	jadi; lahan; ex; trunak; eco; green; park; jaga; lestari; flora; fauna; are; bangkit;
D03	buat; line; backwash; regen; vessel; caton; amon; mached; clearwell
D04	buat; system; instalasi; nafat; air; sisa; backwash; gravity; sand; filter; tuju; terap; proses; reuse; air; bekas; step; water; in; backwash; GSF; siram; tannu; are; PLTU; Paton; terap; proses; recycle; air; step; reuse; siram; GSF; bahan; baku; air; demin;
D05	buat; pipa; salur; air; sisa; dingin; CWP; unit; gravity; sand; filter; WTP
D06	lengkap; baku; baku; standart; listrik; PLTU; Paton
D07	buat; pangung; lengkap; salur; X; mtr; pindah;
D08	Perlu; toilet; lant; SU; toilet; TT; target; bongkar; capa
D09	usul; guna; RFD; bahan; bakar; tambak; cofiring; PLTU; Paton; manfaat; green; power; plat; re; use; sampah; bahan; bakar; PLTU; bahan; lingkun; elola; sampah; tersedia; bahan; bakar; stabil; daya; masyarakat; elola; sampah; RFD
D10	buat; baku; pandu; bac; drawing; diagram; terag
D11	Modifikasi; line; supply; hydrogen; H; plant; unit;
D12	alat; leuba; PMK; damkar; saat; inter; usat; PLTU; Paton;
D13	Management; project; integras; dgn; ellipse
D14	rubah; sistem; dingin; pendam; dgn; pompa; oli; sirkulasi; anoda; cooler; dingin; oli; gas; box; support; aring; hump; dingin; unan; shga; temp; aring; dingin

At the bottom of the table, there is a 'Showing 1 to 14 of 14 entries' label. Below the table, there is a 'Page rendered in 0.5428 seconds. CodeIgniter Version 3.0.0' label. At the bottom right, there is a 'Previous' button, a '1' button, and a 'Next' button.

Gambar 5.5 Antarmuka Stemming

Keterangan Gambar 5.5 bisa dilihat pada Tabel 5.7 di bawah

Tabel 5.7 Antarmuka Stemming

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen.
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya.
<i>Data</i>	Sebagai tempat keseluruhan data tahun 2014.
<i>Tokenizing</i>	Sebagai tempat data yang sudah berupa kata-kata.
<i>Filtering</i>	Sebagai tempat data hasil <i>tokenizing</i> yang sudah dihilangkan <i>stoplist</i> nya.
<i>Stemming</i>	Sebagai tempat data hasil <i>filtering</i> yang sudah dihilangkan kata imbuhan dan akhirnya
TF.IDF.ICSDf	Sebagai tempat kata hasil <i>stemming</i> yang sudah dihitung bobotnya.
<i>Cosine</i>	Sebagai tempat data hasil TF.IDF.ICSDf untuk dihitung kemiripannya dengan <i>query inputan</i> pertama yang selanjutnya digunakan untuk meranking dokumen.
<i>Accurasy</i>	Sebagai tempat data yang sudah dihitung akurasi dengan <i>precision</i> , <i>recall</i> , <i>f-measure</i> dan akurasi

5.4.6 Antarmuka TF.IDF.ICSDf

Query:

Submit

Data

Tokenizing

Filtering

Stemming

TF.IDF.ICSDf

Cosine

Accuracy

Search:

phiw=

phi:

no	Term	Q	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12	D13	D14	df	log(n/df)	wck T	wck NT	Cti T	Cti NT	CSDf	
1	perbanyak	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
2	biopor	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
3	sepanjang	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
4	area	0	2	1	0	1	0	0	0	0	0	0	0	0	0	0	3	0.66900678095858	1	1	0.16666666666667	0.125	0.29166666666667	0.
5	unit	0	1	0	0	0	1	0	0	0	0	0	1	1	0	0	4	0.54406804435028	1	1	0.16666666666667	0.125	0.29166666666667	0.
6	PLTU	1	1	0	0	1	0	1	0	0	2	0	0	1	0	0	5	0.4471583134222	1	1	0.16666666666667	0.125	0.29166666666667	0.
7	hujan	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
8	sumber	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
9	klontong	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
10	serap	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
11	air	0	1	0	0	4	1	0	0	0	0	0	0	0	0	0	3	0.66900678095858	1	1	0.16666666666667	0.125	0.29166666666667	0.
12	hujan	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
13	sekitar	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	0	1	0	0.125	0.125	1.
14	jadi	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	1	0	0.16666666666667	0	0.16666666666667	1.
15	lahan	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	1	0	0.16666666666667	0	0.16666666666667	1.
16	ex	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	1	0	0.16666666666667	0	0.16666666666667	1.
17	trunir	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	1	0	0.16666666666667	0	0.16666666666667	1.
18	eco	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1.1461280356782	1	0	0.16666666666667	0	0.16666666666667	1.

Gambar 5.6 Antarmuka TF.IDF.ICSDf

Keterangan Gambar 5.6 bisa dilihat pada Tabel 5.8 di bawah

Tabel 5.8 Antarmuka TF.IDF.ICSDf

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen.
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya.
<i>Data</i>	Sebagai tempat keseluruhan data tahun 2014.
<i>Tokenizing</i>	Sebagai tempat data yang sudah berupa kata-kata.
<i>Filtering</i>	Sebagai tempat data hasil <i>tokenizing</i> yang sudah dihilangkan <i>stoplist</i> nya.
<i>Stemming</i>	Sebagai tempat data hasil <i>filtering</i> yang sudah dihilangkan kata imbuhan dan akhirnya
TF.IDF.ICSDf	Sebagai tempat kata hasil <i>stemming</i> yang sudah dihitung bobotnya.
<i>Cosine</i>	Sebagai tempat data hasil TF.IDF.ICSDf untuk dihitung kemiripannya dengan <i>query inputan</i> pertama yang selanjutnya digunakan untuk meranking dokumen.
<i>Accurasy</i>	Sebagai tempat data yang sudah dihitung akurasi dengan <i>precision</i> , <i>recall</i> , <i>f-measure</i> dan akurasi

5.4.7 Antarmuka *Cosine Similarity*

Antarmuka *Cosine Similarity* bisa dilihat pada Gambar 5.7 di bawah

The screenshot shows a web application interface for Cosine Similarity. At the top, there is a search bar with the text "Query : pltu" and a "Submit" button. Below the search bar, there are tabs for "Data", "Tokenizing", "Filtering", "Stemming", "TF.IDF.ICSDf", "Cosine", and "Accuracy". The "Cosine" tab is currently selected. The main area displays a table with two columns: "ID" and "Dokumen". The table lists 14 documents (D00 to D14) with their corresponding similarity scores. The scores for D00 to D04 are highlighted in green, while D02 to D14 are in white. The scores range from 0.12104786797319 to 0.048445929724937. At the bottom, there is a pagination bar showing "Showing 1 to 14 of 14 entries" and navigation buttons for "Previous", "1", and "Next".

ID	Dokumen	Hasil Cosine(Di)
D00	Melengkapi buku-buku standar laboratorium untuk PLTU Patun.	0.12104786797319
D12	Mengalihkan limbah lumpur PMK dan air laut bagian di atas unit PLTU patun #12	0.10709523540902
D09	Ditunjuk untuk pengujian RFD sebagai bahan bakar turbin (co-firing) pada PLTU Patun hal ini memberi manfaat sebagai berikut : green power plant : re-use sampah menjadi bahan bakar PLTU, membantu lingkungan dalam mengurangi sampah. : keterbatasan bahan bakar yang stabil, menurunkan emisi gas rumah kaca dalam pengolahan sampah menjadi RFD	0.086327660920016
D01	menyediakan tempat penyimpanan air unit PLTU #12, menyediakan pengaliran baik di atas sumber klontang serta penyempitan air hujan disekitar area #12	0.075129138951799
D04	Pembuatan system instalasi penanaman air dari backwash gravity sand filter, dengan tujuan menerapkan proses reuse air bekas stop water di backwash GSF digunakan untuk penyiraman tanaman di area PLTU Patun, dan menerapkan proses recycle air pada sand stop water dalam kembali ke GSF untuk digunakan sebagai bahan baku air dasar.	0.048445929724937
D02	Menjadikan lahan ex trumix sebagai eco green park sehingga turut menjaga kelestarian flora dan fauna di area sekitar pembangkitan.	0
D03	Pembuatan line ke luar backwash regen dari vessel catton, anion dan mixed menuju clearwell	0
D05	Pembuatan pipa penyaluran air sisa pendingin dari CWP unit #12 menuju gravity sand filter WTP	0
D07	Membuat panggung lengkap ukuran 6 X 8 mtr yang bisa dipindah2:	0
D08	Perlu dibuat toilet di lantai 2 SU sehingga saat ke toilet tidak perlu jauh2 ke TT sehingga target pembongkaran bisa dicapai	0
D10	Membuat buku panduan cara membaca drawing diagram dan keterangannya	0
D11	Modifikasi line supply hydrogen dari H2 plant ke unit #12	0
D13	Management project yang terintegrasi dgn ellipse	0
D14	Merubah sistem pendingin pendingin dgn pompa oli sirkulasi dan penambahan cooler pendingin oli dan gear box support bearingnya diharapkan ada pendingin di pelumasnya sgha temp bearing tetap dingin	0

Gambar 5.7 Antarmuka *Cosine Similarity*

Tabel 5.9 Antarmuka Cosine Similarity

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen.
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya.
<i>Data</i>	Sebagai tempat keseluruhan data tahun 2014.
<i>Tokenizing</i>	Sebagai tempat data yang sudah berupa kata-kata.
<i>Filtering</i>	Sebagai tempat data hasil <i>tokenizing</i> yang sudah dihilangkan <i>stoplist</i> nya.
<i>Stemming</i>	Sebagai tempat data hasil <i>filtering</i> yang sudah dihilangkan kata imbuhan dan akhirnya
TF.IDF.ICSDf	Sebagai tempat kata hasil <i>stemming</i> yang sudah dihitung bobotnya.
<i>Cosine</i>	Sebagai tempat data hasil TF.IDF.ICSDf untuk dihitung kemiripannya dengan <i>query inputan</i> pertama yang selanjutnya digunakan untuk meranking dokumen.
<i>Accuracy</i>	Sebagai tempat data yang sudah dihitung akurasi dengan <i>precision</i> , <i>recall</i> , <i>f-measure</i> dan akurasi

5.4.8 Antarmuka Accuracy

Antarmuka Accuracy relevan bisa dilihat pada Gambar 5.8 di bawah

				Query : pitu	Submit
				phus-phus	
Data Tokenizing Filtering Stemming TFIDF/ICSDf Cosine Accuracy					
No	Dokumen	Relevan	Hasil Cosine	Tidak Relevan	
D06	Melempahi buku-buku standar listrik untuk PLTU Paton	0.12104786797519	☒ Relevan	☐ Tidak Relevan	
D12	Mengaktifkan kembali lembe P&SC dan/atau anatr bagian di sistem unit PLTU paton #12	0.10709523540902	☒ Relevan	☐ Tidak Relevan	
D09	Diusulkan untuk penggunaan RFD sebagai bahan bakar tambahan (cofiring) pada PLTU Paton hal ini memberi manfaat sebagai berikut : green power plant : re-use sampah menjadi bahan bakar PLTU; membantu lingkungan dalam mengelola sampah ; ketersediaan bahan baku yang stabil; memberdayakan masyarakat dalam pengelolaan sampah menjadi RFD	0.086327660920016	☒ Relevan	☐ Tidak Relevan	
D01	menyediakan tempat biopori sepanjang area unit PLTU #12, menyediakan penghijauan baik di atas sumber klontong serta penyerapan air hujan disekitar area #12.	0.075129138951799	☒ Relevan	☐ Tidak Relevan	
D04	Pembuatan system instalasi pemanfaatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse, air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paton dan menerapkan proses recycle, air pada saat step rime diaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.	0.048445929724937	☒ Relevan	☐ Tidak Relevan	
D02	Menjadikan lahan ex trunk sebagai eco green park selangka huri menjaga kelestarian flora dan fauna di area sekitar pembangunan.	0	☐ Relevan	☒ Tidak Relevan	
D03	Pembuatan line keluar backwash regen dari vessel cation, anion dan mixed menuju clearwell	0	☐ Relevan	☒ Tidak Relevan	
D05	Pembuatan pipa penyaluran air sisa pendingin dari CWP unit #12 menuju gravity sand filter WTP	0	☐ Relevan	☒ Tidak Relevan	
D07	Membuat pangung lempahi ukuran 6'X 8' mtr yang bisa dipindah2;	0	☐ Relevan	☒ Tidak Relevan	
D08	Perlu dibuat toilet di lantai 2 SU sehingga saat ke toilet tidak perlu jash2 ke TT sehingga target pembongkaran bisa dicapai	0	☐ Relevan	☒ Tidak Relevan	
D10	Membuat buku panduan cara membaca drawing diagram dan keterangannya	0	☐ Relevan	☒ Tidak Relevan	
D11	Modifikasi line supply hydrogen dari H2 plant ke unit #12	0	☐ Relevan	☒ Tidak Relevan	
D13	Managemen project yang terintegrasi dgn ellipse	0	☐ Relevan	☒ Tidak Relevan	
D14	Merubah sistem pendingin pendingin pompa oli sirkulasi dan penambahan cooler pendingin oli dan gear box support bearingnya diharapkan ada pendingin di pelumasnya slga temp bearing tetap dingin	0	☐ Relevan	☒ Tidak Relevan	
Random				Submit	

Gambar 5.8 Accuracy Relevan

Keterangan Gambar 5.8 bisa dilihat pada Tabel 5.10 di bawah

Tabel 5.10 Antarmuka Accuracy Relevan

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen.
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya.
<i>Data</i>	Sebagai tempat keseluruhan data tahun 2014.
<i>Tokenizing</i>	Sebagai tempat data yang sudah berupa kata-kata.
<i>Filtering</i>	Sebagai tempat data hasil <i>tokenizing</i> yang sudah dihilangkan <i>stoplist</i> nya.
<i>Stemming</i>	Sebagai tempat data hasil <i>filtering</i> yang sudah dihilangkan kata imbuhan dan akhirnya
TF.IDF.ICSDf	Sebagai tempat kata hasil <i>stemming</i> yang sudah dihitung bobotnya.
<i>Cosine</i>	Sebagai tempat data hasil TF.IDF.ICSDf untuk dihitung kemiripannya dengan <i>query inputan</i> pertama yang selanjutnya digunakan untuk meranking dokumen.
<i>Accuracy</i>	Sebagai tempat data yang sudah dihitung akurasi dengan <i>precision</i> , <i>recall</i> , <i>f-measure</i> dan akurasi

Kedua adalah hasil perhitungan akhir dari yang diperoleh dari relevan tidaknya data pada halaman sebelumnya. Antarmuka Accuracy keseluruhan bisa dilihat pada Gambar 5.9 di bawah

Nama	Nilai
TP	5
TN	9
FP	0
FN	0
Recall $(TP / (TP + FN))$	1
Precision $(TP / (TP + FP))$	1
F Measure $(2 * precision * recall / (precision + recall))$	1
F Measure $(2 * TP / ((2 * TP) + FP + FN))$	1
Accuracy $(TP + TN / (TP + TN + FP + FN))$	1

Gambar 5.9 Antarmuka Accuracy Keseluruhan

Keterangan Gambar 5.9 bisa dilihat pada Tabel 5.11 di bawah

Tabel 5.11 Antarmuka Accuracy Keseluruhan

<i>Query</i>	Untuk memasukkan <i>inputan</i> kata untuk mencari dokumen.
<i>Submit</i>	Mengirimkan <i>inputan</i> ke halaman berikutnya.
<i>Data</i>	Sebagai tempat keseluruhan data tahun 2014.
<i>Tokenizing</i>	Sebagai tempat data yang sudah berupa kata-kata.

<i>Filtering</i>	Sebagai tempat data hasil <i>tokenizing</i> yang sudah dihilangkan <i>stoplist</i> nya.
<i>Stemming</i>	Sebagai tempat data hasil <i>filtering</i> yang sudah dihilangkan kata imbuhan dan akhirnya
TF.IDF.ICSDf	Sebagai tempat kata hasil <i>stemming</i> yang sudah dihitung bobotnya.
<i>Cosine</i>	Sebagai tempat data hasil TF.IDF.ICSDf untuk dihitung kemiripannya dengan <i>query inputan</i> pertama yang selanjutnya digunakan untuk meranking dokumen.
<i>Accurasy</i>	Sebagai tempat data yang sudah dihitung akurasi dengan <i>precision, recall, f-measure</i> dan akurasi



BAB 6 PENGUJIAN

Bab ini membahas tentang tahapan pengujian dan analisis dari implementasi metode TF.IDF.ICSDf untuk perankingan dokumen. Pengujian ini menggunakan 207 data dari keseluruhan dokumen ide kreatif PT PJB UP Paiton tahun 2014 yang terdiri dari kelas teknik dan non teknik. Pada penelitian ini akan dilakukan pengujian dengan mencari hasil dari *precision*, *recall* dan *f-measure* menggunakan keseluruhan dokumen dengan dengan rincian masing masing dihitung dengan metode TF.IDF.ICF dan TF.IDF.ICSDf serta menggunakan 3 *query* yang berbeda yang terdiri dari lebih dari 2 kata dan akan dibandingkan dengan *query* tambahan yang mengandung *query* pertama. Selanjutnya hasil *precision*, *recall* dan *f-measure* dari masing-masing metode pada percobaan tersebut akan dibandingkan untuk melihat metode mana yang lebih efektif jika diterapkan pada penelitian ini.

6.1 Hasil dan analisa uji coba

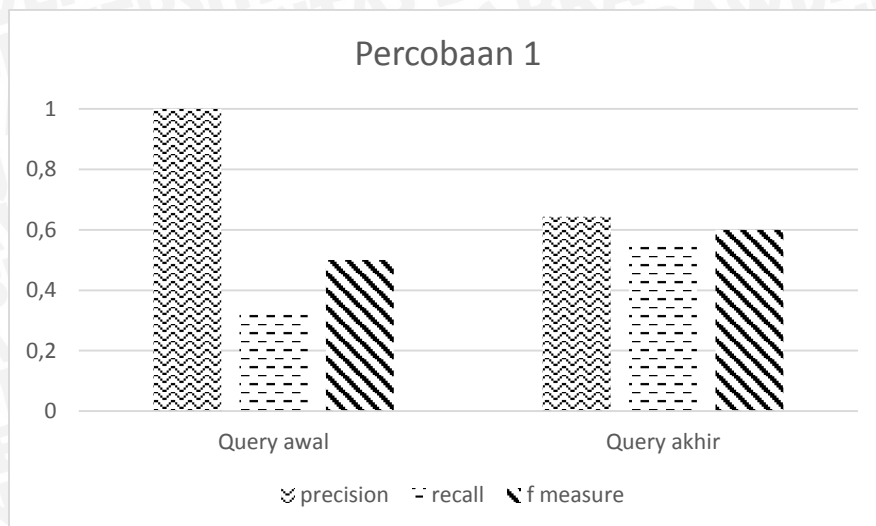
6.1.1 Hasil uji coba pengujian

Pada uji coba ini akan dilakukan pengujian menggunakan keseluruhan data pada tahun 2014 sejumlah 207 data. Terdapat 3 percobaan masing-masing dihitung dengan menggunakan metode TF.IDF.ICF dan TF.IDF.ICSDf dengan menggunakan *query* pertama dan *query* terakhir. Hasil dari *f-measure* dari uji coba ini digunakan untuk perbandingan antara metode TF.IDF dan metode TF.IDF.ICSDf. Hasil *precision*, *recall* dan *f-measure* untuk percobaan 1 *query* awal "team pmk" dan *query* akhir "kriteria perpustakaan team pmk dilengkapi" bisa dilihat pada Tabel 6.1 di bawah ini

Tabel 6.1 TF.IDF.ICSDf precision recall dan f-measure percobaan 1

Indikator	Query Awal	Query Akhir
Precision	1	0,642857
Recall	0,333333	0,5625
F Measure	0,5	0,6

Ilustrasi dalam bentuk *chart* untuk *precision recall* dan *f-measure* percobaan 1 tabel untuk tabel bisa dilihat pada Gambar 6.1 di bawah ini



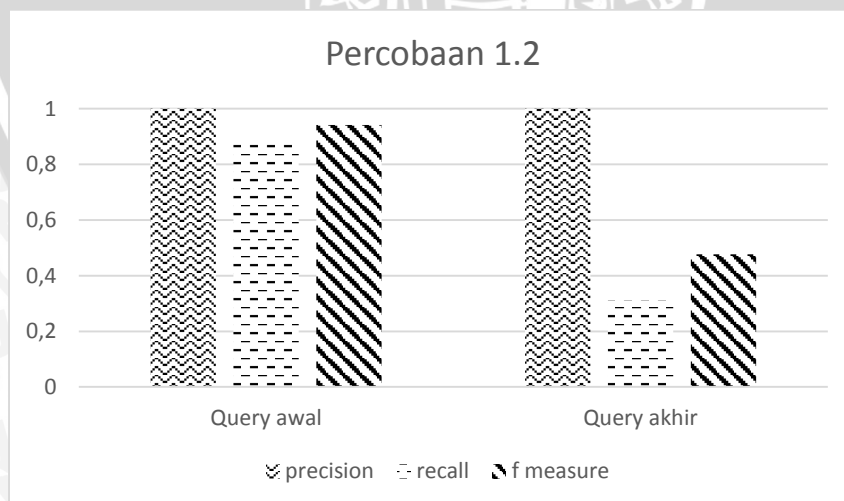
Gambar 6.1 Ilustrasi percobaan 1 pengujian 1

Hasil *precision*, *recall* dan *f-measure* untuk percobaan 1.2 dengan *query* yang sama bisa dilihat pada Tabel 6.2 di bawah ini

Tabel 6.2 TF.IDF.ICF precision recall dan f-measure percobaan 1.2

Indikator	Query Awal	Query Akhir
Precision	1	1
Recall	0,888889	0,3125
F Measure	0,941176	0,47619

Ilustrasi dalam bentuk *chart* untuk *precision recall* dan *f-measure* percobaan 1.2 untuk tabel diatas bisa dilihat pada Gambar 6.2 di bawah ini



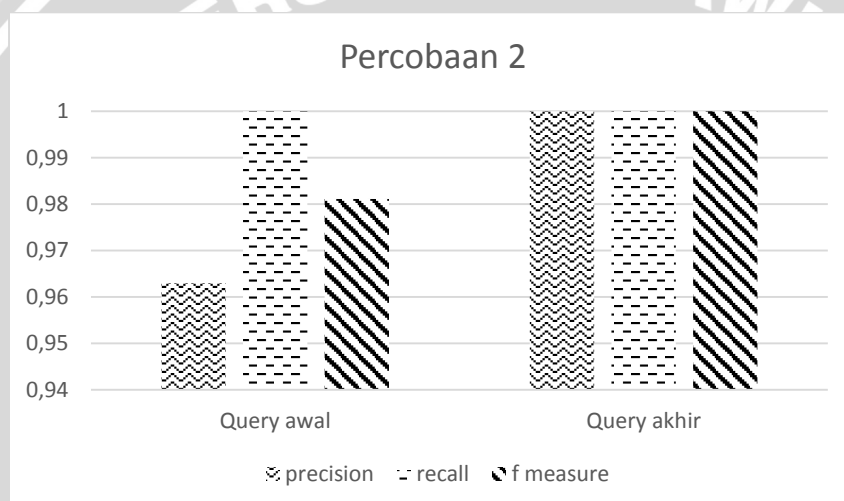
Gambar 6.2 Ilustrasi percobaan 1.2 pengujian 1

Hasil *precision*, *recall* dan *f-measure* untuk percobaan 2 dengan *query* awal “sistem instalasi biodiesel” dan *query* akhir “sistem instalasi biodiesel bagus menarik” bisa dilihat pada Tabel 6.3 di bawah ini

Tabel 6.3 TF.IDF.ICSDf *precision recall* dan *f-measure* percobaan 2

Indikator	Query Awal	Query Akhir
Precision	0,962963	1
Recall	1	1
F Measure	0,981132	1

Ilustrasi dalam bentuk *chart* untuk *precision recall* dan *f-measure* percobaan 2 untuk tabel diatas bisa dilihat pada Gambar 6.3 di bawah ini



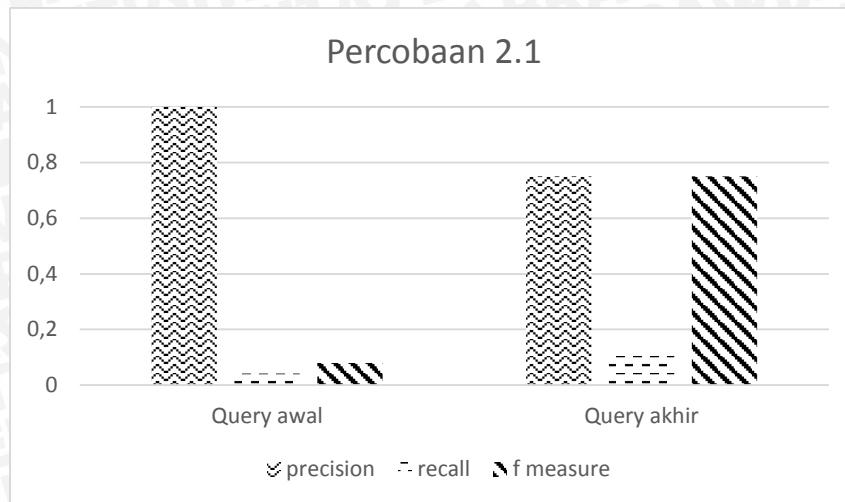
Gambar 6.3 Ilustrasi percobaan 2 pengujian 1

Hasil *precision*, *recall* dan *f-measure* untuk percobaan 2.1 dengan *query* yang sama bisa dilihat pada Tabel 6.4 di bawah ini

Tabel 6.4 TF.IDF.ICF *precision recall* dan *f-measure* percobaan 2.1

Indikator	Query Awal	Query Akhir
Precision	1	0,75
Recall	0,041667	0,115385
F Measure	0,08	0,2

Ilustrasi dalam bentuk *chart* untuk *precision recall* dan *f-measure* percobaan 2.1 untuk tabel diatas bisa dilihat pada Gambar 6.4 di bawah ini



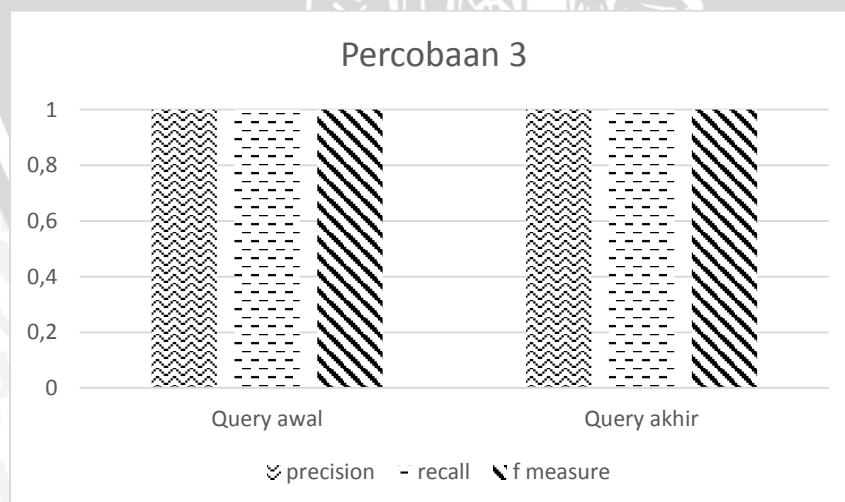
Gambar 6.4 Ilustrasi percobaan 2.1 pengujian 1

Hasil *precision*, *recall* dan *f-measure* untuk percobaan 3 dengan *query* awal “ukuran panggung sempurna” dan *query* akhir “evaluasi ukuran panggung sempurna” bisa dilihat pada Tabel 6.5 di bawah ini

Tabel 6.5 TF.IDF.ICSDf *precision recall* dan *f-measure* percobaan 3

Indikator	Query Awal	Query Akhir
Precision	1	1
Recall	1	1
F Measure	1	1

Ilustrasi dalam bentuk *chart* untuk *precision recall* dan *f-measure* percobaan 3 untuk tabel diatas bisa dilihat pada Gambar 6.5 di bawah ini



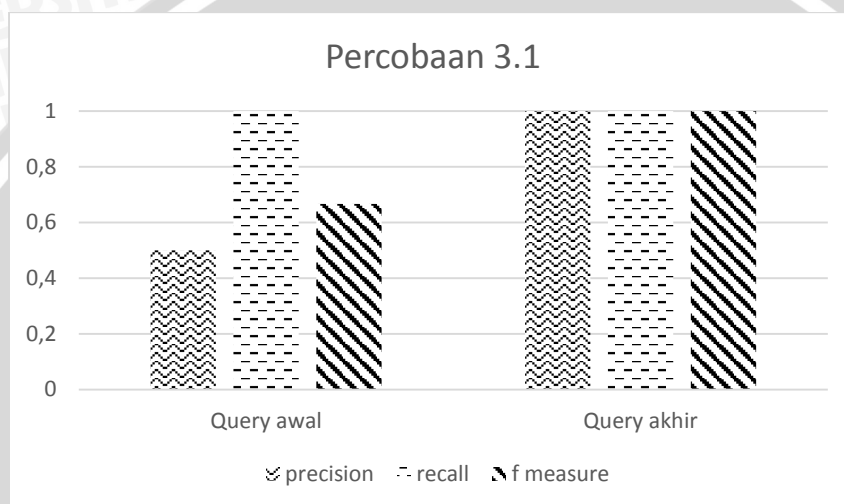
Gambar 6.5 Ilustrasi percobaan 3 pengujian 1

Hasil *precision*, *recall* dan *f-measure* untuk percobaan 3.1 dengan *query* sama bisa dilihat pada Tabel 6.6 di bawah

Tabel 6.6 TF.IDF.ICF *precision recall* dan *f-measure* percobaan 3.1

Indikator	Query Awal	Query Akhir
Precision	1	1
Recall	0,5	0,6
F Measure	0,666667	0,75

Ilustrasi dalam bentuk *chart* untuk *precision recall* dan *f-measure* percobaan 3.1 untuk tabel diatas bisa dilihat pada Gambar 6.6 di bawah ini



Gambar 6.6 Ilustrasi percobaan 3.1 pengujian 1

6.2 Analisa Uji Coba Pengujian

Pada percobaan uji coba diatas didapatkan bahwa hasil untuk *precision*, *recall* dan *f-measure* untuk metode TF.IDF.ICSDf sebagian besar lebih mempunyai nilai lebih tinggi daripada metode TF.IDF.ICF. Ini didapatkan karena saat menggunakan metode TF.IDF.ICF untuk *query* sebagai *inputan* pertama ada pada semua kelas, untuk penelitian ada 2 kelas yaitu Teknik ataupun Nonteknik maka *Invers Class Frequency* dihitung 0 . Metode TF.IDF.ICSDf saat *query* sebagai *inputan* pertama ada pada semua kelas maka tetap dihitung dan terdapat nilai jarak untuk *query* tersebut. Rata-rata hasil dari metode TF.IDF.ICSDf yang telah dihitung dengan *precision*, *recall*, dan *f-measure* untuk *query* awal didapatkan nilai 0,833 begitu pula untuk *query* akhir didapatkan nilai 0,833. Rata-rata metode TF.IDF.ICF yang telah dihitung dengan *precision*, *recall*, dan *f-measure* untuk *query* awal didapatkan nilai 0,526 begitu pula untuk *query* akhir didapatkan nilai 0,433.

Penambahan term pada *query* akhir pada percobaan diatas sebagian besar mempengaruhi kenaikan nilai dari *f-measure*. Ini disebabkan karena *query* semakin spesifik dan jumlah dokumen yang terambil cenderung tetap. Penambahan jumlah kata ini menaikkan ranking dari dokumen yang relevan pada sistem. Semakin banyak kata yang ditambahkan, *query* menjadi semakin spesifik dan dokumen yang diinginkan urutannya akan naik.

BAB 7 KESIMPULAN

Bagian ini memuat kesimpulan dan saran terhadap skripsi. Kesimpulan dan saran disajikan secara terpisah, dengan penjelasan sebagai berikut.

7.1 Kesimpulan

Berdasarkan hasil uji coba parameter metode TF.IDF.ICSDf pada permasalahan perangkingan ide kreatif adalah sebagai berikut

1. Dari percobaan di atas didapatkan hasil bahwa metode TF.IDF.ICSDf dapat bekerja lebih baik daripada metode TF.IDF.ICF Ini diperoleh dari hasil *f-measure* yang lebih tinggi untuk setiap percobaan pada *term* awal ataupun akhir.
2. Algoritma TF.IDF.ICSDf dapat menyelesaikan permasalahan pada perangkingan ide kreatif pada PT PJB UP Paiton yang dihasilkan pada perhitungan *f-measure*. Keberagaman tingkat akurasi diperoleh hasil *f-measure* tertinggi untuk *query* awal yaitu 1 dan terendah yaitu 0,5 sedangkan untuk *query* akhir *f-measure* tertinggi yaitu 1 terkecil yaitu 0,6.

7.2 Saran

Perlu dilakukan penelitian lebih lanjut untuk proses pengembangan perangkingan dokumen dengan menggunakan algoritma *stemming* yang terbaru. Disarankan tidak menggunakan kata yang mengandung kata yang tersimpan pada *filtering* maupun *stemming*.

DAFTAR PUSTAKA

- Azis, Abdillah, Abdul dan Muktyas Indra Bayu. 2012. Implementasi *Vector Space Model* Untuk Pencarian Dokumen. STKIP Surya, Tangerang, Banten.
- Cios, Krzysztof J, dkk. 2007. *Data Mining A Knowledge Discovery Approach*, Springer. (online) <http://link.springer.com/book/10.1007/978-0-387-36795-8> [diakses 26 juli 2016].
- Kusumawardani, Devita. 2009. Temu Kembali Informasi Dengan *Keyword*. Studi Deskriptif Tentang Sistem Temu Kembali Informasi Dengan Controlled Vocabulary Pada Field Judul, Subyek, Dan Pengarang di Perpustakaan Universitas Airlangga. Universitas Airlangga, Surabaya, Jawa Timur.
- Feldman, Ronen dan Sanger, James 2007. *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge University, New York, USA.
- Fuji R., G.S, Mohammad. 2009. *Class-indexing-based term weighting for automatic text classification*, Journal of Informetrics.
- Garcia, Dr. E, 2005. *The Classic Vector Space Model (Description, Advantages and Limitation of the Classic Vector Space Model)*.
- W. J. S. Poerwadarminta, 2007. Kamus Besar Bahasa Indonesia. (online) https://books.google.co.id/books/about/Kamus_umum_Bahasa_Indonesia.html?id=2L9kAAAAMAAJ [diakses 26 juli 2016].
- Hartono, Bambang. 2013. Sistem Informasi Manajemen Berbasis Komputer. Jakarta : Rineka Cipta.
- Jovita, Linda, Andrei H, Derwin S. 2015. *Using Vector Space Model in Answering System*. Ilmu Komputer. Universitas Bina Nusantara.
- Kiftiyani, U. 2014. Perbandingan Algoritma Naïve Bayes dan K Nearest Neighbor untuk Perankingan Dokumen Berbahasa Arab. Program Teknologi Informasi dan Ilmu Komputer.
- Agusta, Ledy. 2009. Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief Dan Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia, Fakultas Teknologi Informasi. Universitas Kristen Satya Wacana.
- Manning, Christopher D., Raghavan, Prabhakar, Schutze. , 2008. Introduction to Information Retrieval. Cambridge University Press, New York, USA.
- Maulana, D. 2014. *Question Answering System Berbasis Clustering* pada Buku Pedoman PTIHK Dengan Algoritma Levenshtein Distance. Program Teknologi Informasi dan Ilmu Komputer.
- Fauzi, Ali, M, Arifin, Zainal, Agus, Yuniarti, Anny, M.Comp.Sc. 2014. *Term Weighting Berbasis Indeks Buku dan Kelas untuk Perankingan Dokumen Berbahasa Arab*. Institut Teknologi Sepuluh Nopember, Surabaya, Jawa timur.

Miswan, 2003. Klasifikasi dan katalogisasi : Sebuah Pengantar. Disampaikan pada "Workshop Perpustakaan dan Kearsipan" yang diselenggarakan oleh STAIN Purwokerto. Tersedia di : www.researchkesos.com/download/jurnal_vol_14.pdf.

Saptari, Janu dan Purwono. 2006. Temu Kembali Informasi Bibliografi Dengan Bahasa Alami Pada Field Judul dan Subyek : Studi Efektifitas Katalog Indik Terpasang Perpustakaan UGM, Berkala Ilmu Perpustakaan dan Informasi – Volume III, Nomor 1 lib.ugm.ac.id/data/pubdata/pusta/janusaptari.pdfugm purwono [diakses 21 September 2011].

Srivastava, Ashok N dan Sahami Mehran. 2009. *Text Mining Classification, Clustering, and Application*. CRC Press, New York.

Tala, Fadillah Z. 2003. A Study OF Stemming Effects on Information Retrieval in Bahasa Indonesia. Master of Logic Project. Institute of Logic, Language and Computation. University van Amsterdam. The Netherland. www.illc.nl/Publication/ResearchReport/Mol-200302.text.pdf [diakses 21 Februari 2016].

Wibowo, Ari. 2012. Peningkatan Performansi Sistem Temu Balik Informasi Dengan Metode Phrasal Translation Dan Query Expansion. Batam: Teknik Multimedia dan Jaringan Politeknik Negeri Batam. Tersedia di : <http://p2m.polibatam.ac.id/wp-content/uploads/2012/05/Ari-WibowoPeningkatan-Performansi-Sistem-Temu-Balik-Informasi.pdf>[diakses 18 Januari 2013].

P, Yoga, Zanwar. 2015. *Query Expansion* Pada Sistem Temu Kembali Informasi Dokumen Berbahasa Indonesia Menggunakan *Pseudo Relevan Feedback*, Fakultas Ilmu Komputer.

LAMPIRAN

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

1.1 Query : team pmk

ID	Kelas	Isi	Relevan	Tidak Relevan
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		
D147	NT	Mengaktifkan kembali lomba PMK/damkar anatr bagian di intern unit PLTU paiton #12		
D145	NT	Dibentuk team pemadam yang anggotanya dari karyawan/rekrutmen tersendiri yang dikhususkan dilakukan setiap minggu 1x latihan agar kesiapan maupun kecakapan dalam menangani kebakaran lebih baik		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		
D006	NT	Memperbanyak spray/memfungsikannya area TH 1 s.d TH5 sehingga apabila loading debunya tidak kemana-mana (polusi). Terutama area TH3 dan TH 5		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D008	NT	memperbanyak tempat biopori disepanjang area unit PLTU #12, memperbanyak penghijauan baik di atas sumber klontong serta penyerapan air hujan disekitar area #12.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

1.2 Query : team pmk

ID	Kelas	Isi	Relevan	Tidak Relevan
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		
D147	NT	Mengaktifkan kembali lomba PMK/damkar anatr bagian di intern unit PLTU paiton #12		
D145	NT	Dibentuk team pemadam yang anggotanya dari karyawan/rekrutmen tersendiri yang dikhususkan dilakukan setiap minggu 1x latihan agar kesiapan maupun kecakapan dalam menangani kebakaran lebih baik		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		
D006	NT	Memperbanyak spray/memfungsikannya area TH 1 s.d TH5 sehingga apabila loading debunya tidak kemana-mana (polusi). Terutama area TH3 dan TH 5		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D008	NT	memperbanyak tempat biopori disepanjang area unit PLTU #12, memperbanyak penghijauan baik di atas sumber klontong serta penyerapan air hujan disekitar area #12.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

1.3 Query : team pmk

ID	Kelas	Isi	Relevan	Tidak Relevan
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		
D147	NT	Mengaktifkan kembali lomba PMK/damkar anatr bagian di intern unit PLTU paiton #12		
D145	NT	Dibentuk team pemadam yang anggotanya dari karyawan/rekrutmen tersendiri yang dikhususkan dilakukan setiap minggu 1x latihan agar kesiapan maupun kecakapan dalam menangani kebakaran lebih baik		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		
D006	NT	Memperbanyak spray/memfungsikannya area TH 1 s.d TH5 sehingga apabila loading debunya tidak kemana-mana (polusi). Terutama area TH3 dan TH 5		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D008	NT	memperbanyak tempat biopori disepanjang area unit PLTU #12, memperbanyak penghijauan baik di atas sumber klontong serta penyerapan air hujan disekitar area #12.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

1.4 Query : kriteria perpustakaan team pmk dilengkapi

ID	Kelas	Isi	Relevan	Tidak Relevan
D107	NT	Cantik tapi tak menarik mungkin itu yang pantas untuk perpustakaan kita UP Paiton, tempatnya bagus ici juga bagus petugasnya juga sudah faknya tapi kenapa tidak menarik? Karena penataannya tidak sesuai dengan kriteria perpustakaan. Buku tentang religi masih campur dgn buku fiksi, buku2nya tidak tertata dgn baik dan rapi		
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		
D147	NT	Mengaktifkan kembali lomba PMK/damkar anatr bagian di intern unit PLTU paiton #12		
D145	NT	Dibentuk team pemadam yang anggotanya dari karyawan/rekrutmen tersendiri yang dikhususkan dilakukan setiap minggu 1x latihan agar kesiapan maupun kecakapan dalam menangani kebakaran lebih baik		
D060	NT	Dibuatkan tempat sampah yang memiliki indikator kwantitas samaph kertas untuk melihat seberapa besar pemakaian kertas dimasing-masing ruangan dan termonitor shg prinsip 3R bisa mendapatkan kriteria proper		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D074	NT	Pembuatan lahan parkir yang strategis dilengkapi dengan atap yg teduh dilengkapi pula peralatan seadanya semisal (pompa sepeda). Dan terakhir perawatan service yang berkala sehingga aset sepeda tetap layak dipergunakan		
D048	NT	Melengkapi buku-buku standart kelistrikan untuk PLTU Paiton		
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D173	NT	dibangun lapangan futsal dan basket diperumahan di lengkapi lampu malam hari,bisa untuk turnamen/laga antar bidang, latihan bisa lebih sering.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

1.5 Query : kriteria perpustakaan team pmk dilengkapi

ID	Kelas	Isi	Relevan	Tidak Relevan
D107	NT	Cantik tapi tak menarik mungkin itu yang pantas untuk perpustakaan kita UP Paiton, tempatnya bagus ici juga bagus petugasnya juga sudah faknya tapi kenapa tidak menarik? Karena penataannya tidak sesuai dengan kriteria perpustakaan. Buku tentang religi masih campur dgn buku fiksi, buku2nya tidak tertata dgn baik dan rapi		
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		
D147	NT	Mengaktifkan kembali lomba PMK/damkar anatr bagian di intern unit PLTU paiton #12		
D145	NT	Dibentuk team pemadam yang anggotanya dari karyawan/rekrutmen tersendiri yang dikhususkan dilakukan setiap minggu 1x latihan agar kesiapan maupun kecakapan dalam menangani kebakaran lebih baik		
D060	NT	Dibuatkan tempat sampah yang memiliki indikator kwantitas samaph kertas untuk melihat seberapa besar pemakaian kertas dimasing-masing ruangan dan termonitor shg prinsip 3R bisa mendapatkan kriteria proper		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D074	NT	Pembuatan lahan parkir yang strategis dilengkapi dengan atap yg teduh dilengkapi pula peralatan seadanya semisal (pompa sepeda). Dan terakhir perawatan service yang berkala sehingga aset sepeda tetap layak dipergunakan		
D048	NT	Melengkapi buku-buku standart kelistrikan untuk PLTU Paiton		
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D173	NT	dibangun lapangan futsal dan basket diperumahan di lengkapi lampu malam hari,bisa untuk turnamen/laga antar bidang, latihan bisa lebih sering.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

1.6 *Query* : kriteria perpustakaan team pmk dilengkapi

ID	Kelas	Isi	Relevan	Tidak Relevan
D107	NT	Cantik tapi tak menarik mungkin itu yang pantas untuk perpustakaan kita UP Paiton, tempatnya bagus ici juga bagus petugasnya juga sudah faknya tapi kenapa tidak menarik? Karena penataannya tidak sesuai dengan kriteria perpustakaan. Buku tentang religi masih campur dgn buku fiksi, buku2nya tidak tertata dgn baik dan rapi		
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		
D147	NT	Mengaktifkan kembali lomba PMK/damkar anatr bagian di intern unit PLTU paiton #12		
D145	NT	Dibentuk team pemadam yang anggotanya dari karyawan/rekrutmen tersendiri yang dikhususkan dilakukan setiap minggu 1x latihan agar kesiapan maupun kecakapan dalam menangani kebakaran lebih baik		
D060	NT	Dibuatkan tempat sampah yang memiliki indikator kwantitas samaph kertas untuk melihat seberapa besar pemakaian kertas dimasing-masing ruangan dan termonitor shg prinsip 3R bisa mendapatkan kriteria proper		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D074	NT	Pembuatan lahan parkir yang strategis dilengkapi dengan atap yg teduh dilengkapi pula peralatan seadanya semisal (pompa sepeda). Dan terakhir perawatan service yang berkala sehingga aset sepeda tetap layak dipergunakan		
D048	NT	Melengkapi buku-buku standart kelistrikan untuk PLTU Paiton		
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D173	NT	dibangun lapangan futsal dan basket diperumahan di lengkapi lampu malam hari,bisa untuk turnamen/laga antar bidang, latihan bisa lebih sering.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

2.1 Query : sistem instalasi biodiesel

ID	Kelas	Isi	Relevan	Tidak Relevan
D114	NT	PLTU Paiton dapat menerapkan cofiring, antara batu bara dan biodiesel dari kemiri sunan ini. PJB dapat bekerjasama dengan masyarakat sekitar maupun PERHUTANI sebagai supplier bahan mentah, sedangkan PJB mendirikan instalasi pengolahannya. Berikut perbandingan kandungan kemiri sunan dengan biodiesel lainnya. keuntungan cofiring tersebut adalah : peningkatan nilai kalor bahan bakar, sehingga efisiensi pembangkitan meningkat, menunjang pencapaian proper, green power plant karena menggunakan bahan bakar terbarukan.		
D019	T	Pembuatan system instalasi pemanfaatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse, air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paiton dan menerapkan proses recycle, air pada saat step rinse disaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D180	T	mengganti sistem analog selector switch 5 channel sengan sistem push botton selector switch terdiri dari atas energize, remote, local		
D115	T	Memodifikasi pemanas yang ada di hot water dengan mengadopsi sistem oven		
D120	T	Modifikasi sistem thermal drain untuk menjada kualitas steam sootblower		
D205	T	usulan dipasang AWBS yg sudah teruji pd beberapa pembangkit di dunia. Prinsip kerja sistem ini adalah mendeteksi suhu yg terdeteksi oleh sensor yg telah ditentukan sehingga dpt diproses secara menyeluruh untuk kemudian dikalkulasi sehingga dapat		

		ditentukan oleh sistem, sootblower mana yang akan dioperasikan secara efisien		
D131	NT	Dengan menggunakan helikopter remote kamera bisa memantau kondisi peralatan secara remote/mobile dan tentunya sistem keamanan lebih handal		
D136	T	Pemasangan lampu penerangan dengan menggunakan sistem panel surya/solar cell. Hal ini dapat mengurangi pemakaian listrik sehingga menambah efisiensi pembangkit		



Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

2.2 Query : sistem instalasi biodiesel

ID	Kelas	Isi	Relevan	Tidak Relevan
D114	NT	PLTU Paiton dapat menerapkan cofiring, antara batu bara dan biodiesel dari kemiri sunan ini. PJB dapat bekerjasama dengan masyarakat sekitar maupun PERHUTANI sebagai supplier bahan mentah, sedangkan PJB mendirikan instalasi pengolahannya. Berikut perbandingan kandungan kemiri sunan dengan biodiesel lainnya. keuntungan cofiring tersebut adalah : peningkatan nilai kalor bahan bakar, sehingga efisiensi pembangkitan meningkat, menunjang pencapaian proper, green power plant karena menggunakan bahan bakar terbarukan.		
D019	T	Pembuatan system instalasi pemanfaatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse, air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paiton dan menerapkan proses recycle, air pada saat step rinse disaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D180	T	mengganti sistem analog selector switch 5 channel sengan sistem push botton selector switch terdiri dari atas energize, remote, local		
D115	T	Memodifikasi pemanas yang ada di hot water dengan mengadopsi sistem oven		
D120	T	Modifikasi sistem thermal drain untuk menjada kualitas steam sootblower		
D205	T	usulan dipasang AWBS yg sudah teruji pd beberapa pembangkit di dunia. Prinsip kerja sistem ini adalah mendeteksi suhu yg terdeteksi oleh sensor yg telah ditentukan sehingga dpt diproses secara menyeluruh untuk kemudian dikalkulasi sehingga dapat		

		ditentukan oleh sistem, sootblower mana yang akan dioperasikan secara efisien		
D131	NT	Dengan menggunakan helikopter remote kamera bisa memantau kondisi peralatan secara remote/mobile dan tentunya sistem keamanan lebih handal		
D136	T	Pemasangan lampu penerangan dengan menggunakan sistem panel surya/solar cell. Hal ini dapat mengurangi pemakaian listrik sehingga menambah efisiensi pembangkit		



Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

2.3 Query : sistem instalasi biodiesel

ID	Kelas	Isi	Relevan	Tidak Relevan
D114	NT	PLTU Paiton dapat menerapkan cofiring, antara batu bara dan biodiesel dari kemiri sunan ini. PJB dapat bekerjasama dengan masyarakat sekitar maupun PERHUTANI sebagai supplier bahan mentah, sedangkan PJB mendirikan instalasi pengolahannya. Berikut perbandingan kandungan kemiri sunan dengan biodiesel lainnya. keuntungan cofiring tersebut adalah : peningkatan nilai kalor bahan bakar, sehingga efisiensi pembangkitan meningkat, menunjang pencapaian proper, green power plant karena menggunakan bahan bakar terbarukan.		
D019	T	Pembuatan system instalasi pemanfaatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse, air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paiton dan menerapkan proses recycle, air pada saat step rinse disaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D180	T	mengganti sistem analog selector switch 5 channel sengan sistem push botton selector switch terdiri dari atas energize, remote, local		
D115	T	Memodifikasi pemanas yang ada di hot water dengan mengadopsi sistem oven		
D120	T	Modifikasi sistem thermal drain untuk menjada kualitas steam sootblower		
D205	T	usulan dipasang AWBS yg sudah teruji pd beberapa pembangkit di dunia. Prinsip kerja sistem ini adalah mendeteksi suhu yg terdeteksi oleh sensor yg telah ditentukan sehingga dpt diproses secara menyeluruh untuk kemudian dikalkulasi sehingga dapat		

		ditentukan oleh sistem, sootblower mana yang akan dioperasikan secara efisien		
D131	NT	Dengan menggunakan helikopter remote kamera bisa memantau kondisi peralatan secara remote/mobile dan tentunya sistem keamanan lebih handal		
D136	T	Pemasangan lampu penerangan dengan menggunakan sistem panel surya/solar cell. Hal ini dapat mengurangi pemakaian listrik sehingga menambah efisiensi pembangkit		



Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

2.4 *Query* : sistem instalasi biodiesel bagus menarik

ID	Kelas	Isi	Relevan	Tidak Relevan
D107	NT	Cantik tapi tak menarik mungkin itu yang pantas untuk perpustakaan kita UP Paiton, tempatnya bagus ici juga bagus petugasnya juga sudah faknya tapi kenapa tidak menarik? Karena penataannya tidak sesuai dengan kriteria perpustakaan. Buku tentang religi masih campur dgn buku fiksi, buku2nya tidak tertata dgn baik dan rapi		
D114	NT	PLTU Paiton dapat menerapkan cofiring, antara batu bara dan biodiesel dari kemiri sunan ini. PJB dapat bekerjasama dengan masyarakat sekitar maupun PERHUTANI sebagai supplier bahan mentah, sedangkan PJB mendirikan instalasi pengolahannya. Berikut perbandingan kandungan kemiri sunan dengan biodiesel lainnya. keuntungan confiring tersebut adalah : peningkatan nilai kalor bahan bakar, sehingga efisiensi pembangkitan meningkat, menunjang pencapaian proper, green power plant karena menggunakan bahan bakar terbarukan.		
D089	NT	buku berjudul paiton membingkai alam sederhana tapi menarik, alangkah baiknya jika buku yg ditulis oleh TIM Fakultas kehutanan IPB itu bisa dilipat gandakan dan dijual utk masyarakat umum. Karena selain yg isinya mengandung ilmu pengetahuannya gambarnya begitu indah ,cerah dan jelas shga sangat bagus untuk dikonsumsi anak-anak utk mengenalkan mereka pd keanekaragaman hayati		
D019	T	Pembuatan system instalasi pemanfaatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse,air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paiton dan menerapkan proses recycle, air pada saat step rinse disaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.		
D096	T	Solusinya mengganti penggerak valve butterfly tersebut dengan jenis penggerak elektrik modul auma karena model ini handal dam presisi kabel demand 4 - 20 mA sudah tersedia. Hanya tinggal tarik kabel 300V		

		dari breaker yg letaknya cukup dekat dan membuat supportnya		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D180	T	mengganti sistem analog selector switch 5 channel sengan sistem push botton selector switch terdiri dari atas energize, remote, local		
D155	T	Memodifikasi pemanas yang ada di hot water dengan mengadopsi sistem oven		
D120	T	Modifikasi sistem thermal drain untuk menjada kualitas steam sootblower		



Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

2.5 Query : sistem instalasi biodiesel bagus menarik

ID	Kelas	Isi	Relevan	Tidak Relevan
D107	NT	Cantik tapi tak menarik mungkin itu yang pantas untuk perpustakaan kita UP Paiton, tempatnya bagus ici juga bagus petugasnya juga sudah faknya tapi kenapa tidak menarik? Karena penataannya tidak sesuai dengan kriteria perpustakaan. Buku tentang religi masih campur dgn buku fiksi, buku2nya tidak tertata dgn baik dan rapi		
D114	NT	PLTU Paiton dapat menerapkan cofiring, antara batu bara dan biodiesel dari kemiri sunan ini. PJB dapat bekerjasama dengan masyarakat sekitar maupun PERHUTANI sebagai supplier bahan mentah, sedangkan PJB mendirikan instalasi pengolahannya. Berikut perbandingan kandungan kemiri sunan dengan biodiesel lainnya. keuntungan confiring tersebut adalah : peningkatan nilai kalor bahan bakar, sehingga efisiensi pembangkitan meningkat, menunjang pencapaian proper, green power plant karena menggunakan bahan bakar terbarukan.		
D089	NT	buku berjudul paiton membingkai alam sederhana tapi menarik, alangkah baiknya jika buku yg ditulis oleh TIM Fakultas kehutanan IPB itu bisa dilipat gandakan dan dijual utk masyarakat umum. Karena selain yg isinya mengandung ilmu pengetahuannya gambarnya begitu indah ,cerah dan jelas shga sangat bagus untuk dikonsumsi anak-anak utk mengenalkan mereka pd keanekaragaman hayati		
D019	T	Pembuatan system instalasi pemanfaatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse,air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paiton dan menerapkan proses recycle, air pada saat step rinse disaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.		
D096	T	Solusinya mengganti penggerak valve butterfly tersebut dengan jenis penggerak elektrik modul auma karena model ini handal dam presisi kabel demand 4 - 20 mA sudah tersedia. Hanya tinggal tarik kabel 300V		

		dari breaker yg letaknya cukup dekat dan membuat supportnya		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D180	T	mengganti sistem analog selector switch 5 channel sengan sistem push botton selector switch terdiri dari atas energize, remote, local		
D155	T	Memodifikasi pemanas yang ada di hot water dengan mengadopsi sistem oven		
D120	T	Modifikasi sistem thermal drain untuk menjada kualitas steam sootblower		



Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

2.6 Query : sistem instalasi biodiesel bagus menarik

ID	Kelas	Isi	Relevan	Tidak Relevan
D107	NT	Cantik tapi tak menarik mungkin itu yang pantas untuk perpustakaan kita UP Paiton, tempatnya bagus ici juga bagus petugasnya juga sudah faknya tapi kenapa tidak menarik? Karena penataannya tidak sesuai dengan kriteria perpustakaan. Buku tentang religi masih campur dgn buku fiksi, buku2nya tidak tertata dgn baik dan rapi		
D114	NT	PLTU Paiton dapat menerapkan cofiring, antara batu bara dan biodiesel dari kemiri sunan ini. PJB dapat bekerjasama dengan masyarakat sekitar maupun PERHUTANI sebagai supplier bahan mentah, sedangkan PJB mendirikan instalasi pengolahannya. Berikut perbandingan kandungan kemiri sunan dengan biodiesel lainnya. keuntungan confiring tersebut adalah : peningkatan nilai kalor bahan bakar, sehingga efisiensi pembangkitan meningkat, menunjang pencapaian proper, green power plant karena menggunakan bahan bakar terbarukan.		
D089	NT	buku berjudul paiton membingkai alam sederhana tapi menarik, alangkah baiknya jika buku yg ditulis oleh TIM Fakultas kehutanan IPB itu bisa dilipat gandakan dan dijual utk masyarakat umum. Karena selain yg isinya mengandung ilmu pengetahuannya gambarnya begitu indah ,cerah dan jelas shga sangat bagus untuk dikonsumsi anak-anak utk mengenalkan mereka pd keanekaragaman hayati		
D019	T	Pembuatan system instalasi pemanfaatan air sisa backwash gravity sand filter, dengan tujuan menerapkan proses reuse,air bekas step water in backwash GSF digunakan untuk penyiraman tanaman di area PLTU Paiton dan menerapkan proses recycle, air pada saat step rinse disaring kembali ke GSF untuk digunakan sebagai bahan baku air demin.		
D096	T	Solusinya mengganti penggerak valve butterfly tersebut dengan jenis penggerak elektrik modul auma karena model ini handal dam presisi kabel demand 4 - 20 mA sudah tersedia. Hanya tinggal tarik kabel 300V		

		dari breaker yg letaknya cukup dekat dan membuat supportnya		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D180	T	mengganti sistem analog selector switch 5 channel sengan sistem push botton selector switch terdiri dari atas energize, remote, local		
D155	T	Memodifikasi pemanas yang ada di hot water dengan mengadopsi sistem oven		
D120	T	Modifikasi sistem thermal drain untuk menjada kualitas steam sootblower		



Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

3.1 Query : ukuran panggung sempurna

ID	Kelas	Isi	Relevan	Tidak Relevan
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D156	NT	Dipasang sparator pada pegangan tanga dengan ukuran 30 s/d 40 cm di sepanjang pegangan tanga khususnya tanga yang ada di SU		
D062	T	Pengadaan coal nozzle 1 buah utk acuan reverse engineering sehingga diharapkan tercapai sistem pembakaran yg sempurna diboiler		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

3.2 Query : ukuran panggung sempurna

ID	Kelas	Isi	Relevan	Tidak Relevan
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D156	NT	Dipasang sparator pada pegangan tanga dengan ukuran 30 s/d 40 cm di sepanjang pegangan tanga khususnya tanga yang ada di SU		
D062	T	Pengadaan coal nozzle 1 buah utk acuan reverse engineering sehingga diharapkan tercapai sistem pembakaran yg sempurna diboiler		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

3.3 Query : ukuran panggung sempurna

ID	Kelas	Isi	Relevan	Tidak Relevan
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D156	NT	Dipasang sparator pada pegangan tanga dengan ukuran 30 s/d 40 cm di sepanjang pegangan tanga khususnya tanga yang ada di SU		
D062	T	Pengadaan coal nozzle 1 buah utk acuan reverse engineering sehingga diharapkan tercapai sistem pembakaran yg sempurna diboiler		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		
D005	NT	Secepatnya kita memiliki team PMK yang minimal satu regu tiga personel (1 driver, 2 anggota untuk tiap ship) karena bahaya kebakaran sewaktu-waktu bisa terjadi.		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

3.4 Query : evaluasi ukuran panggung sempurna

ID	Kelas	Isi	Relevan	Tidak Relevan
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D111	T	Dilakukan evaluasi drawing/diagram kelistrikan agar disesuaikan dengan kondisi sebenarnya.		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D156	NT	Dipasang sparator pada pegangan tanga dengan ukuran 30 s/d 40 cm di sepanjang pegangan tanga khususnya tanga yang ada di SU		
D062	T	Pengadaan coal nozzle 1 buah utk acuan reverse engineering sehingga diharapkan tercapai sistem pembakaran yg sempurna diboiler		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

3.5 Query : evaluasi ukuran panggung sempurna

ID	Kelas	Isi	Relevan	Tidak Relevan
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D111	T	Dilakukan evaluasi drawing/diagram kelistrikan agar disesuaikan dengan kondisi sebenarnya.		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D156	NT	Dipasang sparator pada pegangan tanga dengan ukuran 30 s/d 40 cm di sepanjang pegangan tanga khususnya tanga yang ada di SU		
D062	T	Pengadaan coal nozzle 1 buah utk acuan reverse engineering sehingga diharapkan tercapai sistem pembakaran yg sempurna diboiler		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		

Nama, Umur :

Tanggal isi :

Berikan centang (v) sesuai pilihan anda pada kolom relevan dan tidak relevan sesuai kemiripan *query* pada setiap dokumen yang telah disediakan.

3.6 Query : evaluasi ukuran panggung sempurna

ID	Kelas	Isi	Relevan	Tidak Relevan
D054	NT	Membuat panggung lengkapi ukuran 6 X 8 mtr yang bisa dipindah2;		
D111	T	Dilakukan evaluasi drawing/diagram kelistrikan agar disesuaikan dengan kondisi sebenarnya.		
D007	NT	menyempurnakan & melengkapi bin lokasi gudang		
D156	NT	Dipasang sparator pada pegangan tanga dengan ukuran 30 s/d 40 cm di sepanjang pegangan tanga khususnya tanga yang ada di SU		
D062	T	Pengadaan coal nozzle 1 buah utk acuan reverse engineering sehingga diharapkan tercapai sistem pembakaran yg sempurna diboiler		
D063	T	Penambahan coller pendingin boiler untuk sistem pelumasan di gear box support bearing SAH diharapkan sistem pelumasannya sempurna		
D001	T	Optimalisasi/upgrade level detector di hopper ESP supaya dapat mengontrol akurat dan terintegrasi dengan control sistem di ESP dan fly ash sistem.		
D002	T	Modifikasi wiring penthouse heater dan blower serta mengintegrasikannya dengan modul pada EPIC III yang sudah ada.		
D003	T	Pengantian tutup/cover kopling ERM pada ESP dengan menggunakan material yang tembus cahaya seperti kaca akrilik.		
D004	T	dibuatkan tempat penampungan limbah sementara, sehingga area utara labor bersih dan rapi		

























