

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

Riza Aprilia Sari<sup>1)</sup>, Indriati<sup>2)</sup>, Rizal Setya Perdana<sup>3)</sup>

Program Studi Informatika, Fakultas Ilmu Komputer  
Universitas Brawijaya

Jl. Veteran, Malang 65145, Indonesia

Email: rizaaprilial2[at]gmail.com<sup>1)</sup>, indriati.tif[at]ub.a.c.id<sup>2)</sup>, rizalespe[at]ub.ac.id<sup>3)</sup>

**Abstrak**

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. Karena itu dibuatlah program pencarian kata kunci 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 *cosin similarity* digunakan untuk meranking dokumen . Perolehan hasil perankingan selanjutnya digunakan untuk menghitung akurasi dengan memilih dokumen yang relevan. Metode *Inverse Class Space Density Frequency* digunakan untuk pembobotan hasil dari dokumen relevan sesuai kata kunci yang dimasukkan dengan rata-rata akurasi *query* awal 0,667 dan rata-rata *query* akhir 0,833.

**Kata kunci:** *Inverse Class Space Density Frequency, Cosine Similarity*

**1. PENDAHULUAN**

**1.1. Latar Belakang**

Dokumen merupakan setiap benda yang memuat atau berisi rekaman informasi yang harus ditata dan dikelola dengan baik. Dokumen-dokumen yang dimiliki oleh Instansi baik swasta maupun negeri makin lama makin banyak, dan memerlukan tempat untuk menyimpan dokumen-dokumen tersebut. Informasi adalah data yang diolah menjadi bentuk yang dapat dijadikan dasar mengambil keputusan yang tepat (Taryana Suryana, 2012).

Dengan semakin banyaknya kebutuhan akan Infomasi 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.

Metode *Inverse Class Space Density Frequency* (ICS<sub>d</sub>F) merupakan salah satu metode pembobotan kata yang cocok digunakan untuk perankingan dokumen dengan menggunakan kata kunci.

Penelitian terkait yang menggunakan metode ini adalah penelitian yang berjudul "*Class indexing based term weighting for automatic text classification*" oleh Fuji Ren dan Mohammad Golam Sohrab, 2013. Disini dijelaskan bahwa metode *invers class space density frequency* memiliki *precision* dan *recall* yang lebih tinggi daripada metode TF.IDF.ICF saat

diimplementasikan pada klasifikasi data dengan menggunakan metode *Naive Bayes*.

Dokumen yang akan digunakan pada penelitian ini adalah data Ide Kreatif yang ada pada Perusahaan PJB UP Paiton. Data Ide Kreatif merupakan kalimat-kalimat Ide atau saran berbahasa Indonesia untuk selanjutnya dipakai sebagai peningkatan pembangunan perusahaan yang diperoleh dari *event* pengumpulan kritik dan saran pada perusahaan tersebut yang diadakan setiap tahun. Sistem pencari ini berguna untuk menemukan data pada dokumen ide kreatif yang tersimpan pada *database*. Mesin pencari menggunakan *query* / 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<sub>d</sub>F) 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, Mohammad, 2013).

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<sub>s</sub>F)”.

### 1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan di atas maka dapat disimpulkan rumusan masalah sebagai berikut :

1. Apakah hasil uji coba penerapan metode *Inverse Space Density Frequency* dalam Sistem Temu Kembali Informasi untuk dokumen Ide Kreatif ?
2. Apakah keberagaman tingkat akurasi hasil perhitungan dengan metode *Invers Space Density Frequency* dalam pencarian dokumen Ide Kreatif ?

### 1.3. Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini yaitu :

1. Input berupa *query* atau kata kunci berupa teks..
2. Dokumen yang digunakan adalah dokumen Ide Kreatif PJB UP Paiton tahun 2014.
3. Output berupa perankingan data dari dokumen yang sesuai dengan kata kunci yang dimasukkan.
3. Mayoritas dokumen menggunakan bahasa Indonesia.

### 1.4. Tujuan

Berdasarkan rumusan masalah yang dikemukakan, tujuan yang ingin penulis capai dalam penelitian ini yaitu :

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

## 2. LANDASAN TEORI

### 2.1 Sistem Temu Kembali Informasi

Secara sederhana temu kembali informasi merupakan suatu sistem yang menyimpan informasi dan menemukan kembali informasi tersebut (Janu Suptari; 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 (Lauren B. Doyle 2003)

### 2.2 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 (Garcia, 2005)

### 2.3 Tokenizing

Tahap *tokenizing* adalah tahap pemotongan *string input* berdasarkan tiap kata yang menyusunnya . 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*. (Garcia, 2005)

### 2.4 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 kepunyaan. Memperhitungkan *stopword* pada transformasi teks akan membuat keseluruhan sistem *text mining* bergantung kepada faktor bahasa. Hal ini menjadi kelemahan dari proses penghilangan *stopword* (Tala, Fadillah, 2003).

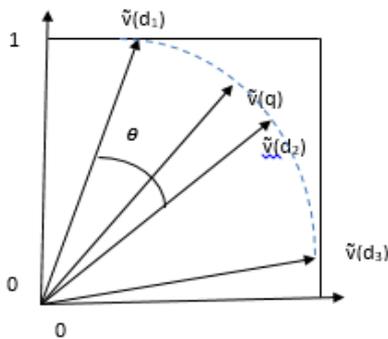
### 2.5 Stemming Porter Algorithm

*Stemming* adalah suatu proses yang menyediakan suatu pemetaan antara berbagi kata dengan morfologi yang berbeda menjadi satu bentuk dasar (*stem*) (Tala, 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

### 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 1 Representasi Perumusan Cosine Similarity

Dalam Gambar 1 terdapat tiga vektor dokumen d1, d2 dan d3 dan satu vektor query q. *Cosine similarity* menghitung nilai kosinus  $\theta$  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. kosine dinyatakan sebagai berikut :

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

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* optimal.

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  $t$  *term* yang diekstrak dari *database*, maka nilai *cosine similarity* didefinisikan sebagai berikut :

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

Keterangan :

- $|q|$  : normalisasi *query*
- $|d_j|$  : normalisasi dokumen
- $W_{iq}$  : bobot *term i* pada *query*
- $W_{ij}$  : bobot *term i* pada *query*

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, 2015)

### 2.7 Inverse Class Space Density Frequency

*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  $TF.IDF.ICSD_F$  untuk *term*  $t_i$  pada dokumen  $d_j$  dengan kategori  $c_k$  didefinisikan sebagai berikut :

Keterangan :

- $W_{TF.IDF.ICSD_F}$  : 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 :

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

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<sub>s</sub>F). Dengan ICF dan ICS<sub>s</sub>F ini *term* yang sering muncul pada banyak kelas akan memiliki nilai yang kecil. (Fuji Ren, Mohammad, 2009).

**2.8 Precision, Recall, 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 standart 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 2.1 Confusion Matrix**

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

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) :

$$precision = \frac{TP}{TP+FP} \tag{2.10}$$

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

$$recall = \frac{TP}{TP+FN} \tag{2.11}$$

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

$$akurasi = \frac{TP+TN}{TP+FP+TN+FN} \tag{2.12}$$

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 :

$$F1 = \frac{2 \times recall \times precision}{recall+precision} \tag{2.13}$$

**3. METODOLOGI PENELITIAN**

Demi terselesaikannya 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*. Tahapan-tahapan dalam penelitian tersebut dapat diilustrasikan dengan diagram blok metodologi penelitian seperti pada Gambar 3.1 dibawah ini :





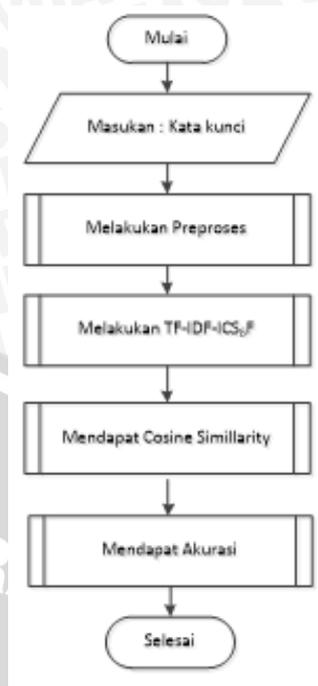
**Gambar 2** Metodologi Penelitian

Berdasarkan gambar 2 tahapan-tahapan yang dilakukan dalam penelitian ini adalah :

1. Melakukan studi literatur mengenai algoritma TF.IDF.ICSDf, data yang digunakan yaitu data ide kreatif pada PJB UP Paiton, dan literatur lain terkait metode dan masalah yang digunakan.
2. Pengumpulan dataset yang digunakan, yaitu data ide kreatif pada PT PJB UP Paiton pada tahun 2014.
3. Melakukan perancangan sistem menggunakan TF.IDF.ICSDf untuk melakukan pembobotan pada tiap kata serta perankingan untuk dokumen.
4. Mengimplementasikan hasil analisis dan perancangan yang telah dilakukan dalam bentuk perangkat lunak untuk pembobotan dan perankingan data ide kreatif yang telah ditentukan menggunakan TF.IDF.ICSDf.
5. Melakukan pengujian dan analisis terhadap perangkat lunak.
6. Proses pengambilan keputusan untuk menyimpulkan keseluruhan mengenai penelitian yang dilakukan.

#### 4. PERANCANGAN

Bab ini membahas bagaimana gambaran program yang akan dibuat sesuai dengan bahasan yang ada pada bab pendahuluan. Yaitu 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.ICSDf selanjutnya dihitung dengan *cosine similarity* untuk dihitung jarak kemiripan antar dokumen sebelum akhirnya dihitung keakurasiannya. Flowchart sistem secara umum bisa dilihat pada Gambar 3 dibawah ini :



**Gambar 3** Flowchart Sistem

Berdasarkan gambar 3, proses pada sistem meliputi 5 proses yaitu:

1. Memasukkan kata kunci sebagai langkah pertama mencari dokumen
2. Data mentah akan diolah dengan tahap *preprocessing* yaitu meliputi *tokenizing, filtering, stemming*.
3. Hasil data tahap *preprocessing* dihitung dengan metode TF.IDF.ICSDf.
4. Hasil perhitungan TF.IDF.ICSDf dicari kemiripan datanya dengan *Cosine Similarity* yang selanjutnya digunakan untuk meranking dokumen.
5. Hasil perankingan akan digunakan untuk menentukan dokumen relevan sebagai acuan menghitung akurasi.

#### 5. IMPLEMENTASI

Pada bab ini menjelaskan tampilan awal antarmuka sistem, hasil dari proses *cosine similarity* dan akurasi bisa dilihat dibawah ini :

Query :

**Gambar 4** Implementasi Antarmuka Beranda

Gambar 4 diatas merupakan tampilan awal sistem sebagai tempat memasukkan kata kunci.

**Tabel 5.1** Hasil Cosine

ID	Dokumen	Cosine
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.	0.3314 765151 7895
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.	0.0825 324803 3451

Tabel 5.1 diatas merupakan hasil dari *cosine similarity* pada program yang digunakan sebagai perankingan dari *query* sebelumnya.

**Tabel 5.2** Hasil Akurasi

Nama	Nilai
TP	1
TN	205
FP	1
FN	0
Recall $\{TP / (TP + FP)\}$	0.5
Precision $\{TP / (TP + FN)\}$	1
F Measure $\{2 * precision * recall / (precision + recall)\}$	0.666666666666667
F Measure $\{2 * TP / ((2 * TP) + FP + FN)\}$	0.666666666666667
Accuracy $\{TP * TN / (TP + TN + FP + FN)\}$	0.666666666666667

Tabel 5.2 diatas merupakan hasil akurasi dari program untuk *query* sebelumnya.

## 6. PENGUJIAN DAN ANALISIS

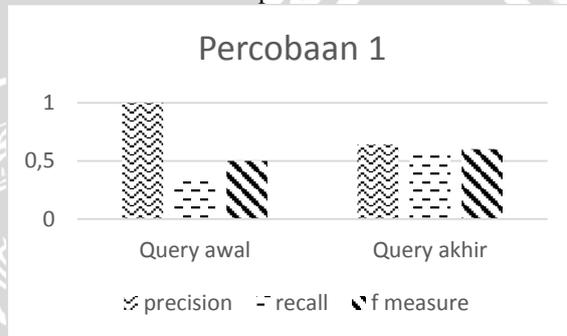
Bab ini membahas tentang tahapan pengujian dan analisis dari implementasi metode TF.IDF.ICSDf untuk perankingan dokumen. Pengujian ini menggunakan 207 data dari keseluruhan dokumen ide kreatif PT PJB UP Paiton tahun 2014 yang terdiri dari kelas teknik dan nonteknik. Terdapat dua jenis pengujian. Pengujian pertama 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 . Pengujian kedua yaitu membandingkan keefektifitasan metode TF.IDF.ICF dan TF.IDF.ICSDf dengan hasil f-measure dari masing-masing metode

### 6.1 Hasil dan Analisa uji coba pengujian 1

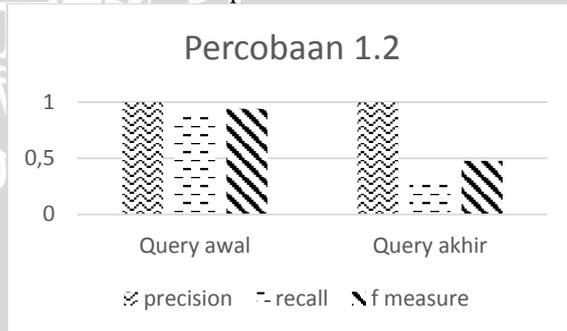
#### 6.1.1 Hasil dan Analisa uji coba percobaan 1

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* akhir. Grafik hasil *precision*, *recall* dan *f-measure* untuk percobaan 1 *query* awal “team pmk” dan *query* akhir ”kriteria perpustakaan team pmk dilengkapi” untuk metode TF.IDF.ICSDf bisa dilihat pada Gambar 5 dibawah ini :



**Gambar 5** Grafik percobaan 1 TF.IDF.ICSDf

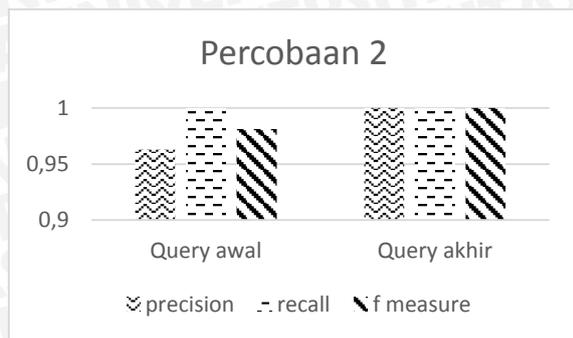
Grafik hasil *precision*, *recall* dan *f-measure* untuk percobaan 1.2 dengan *query* awal “team pmk” dan *query* akhir ”kriteria perpustakaan team pmk dilengkapi” untuk metode TF.IDF.ICF bisa dilihat pada Gambar 6 dibawah ini :



**Gambar 6** Grafik percobaan 1.2 TF.IDF.ICF

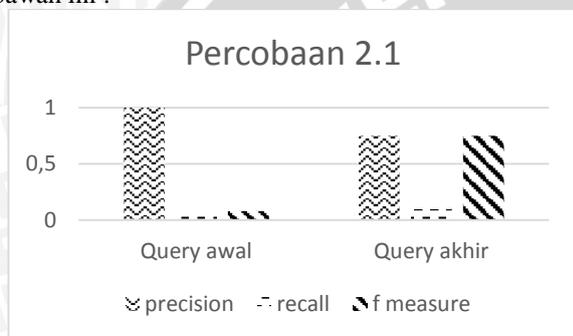
#### 6.1.2 Hasil dan Analisa uji coba percobaan 2

Grafik hasil *precision*, *recall* dan *f-measure* untuk percobaan 2 dengan *query* awal “sistem instalasi biodiesel” dan *query* akhir “sistem instalasi biodiesel bagus menarik” untuk metode TF.IDF.ICSDf bisa dilihat pada Gambar 7 dibawah ini :



Gambar 7 Grafik percobaan 2 TF.IDF.ICSdF

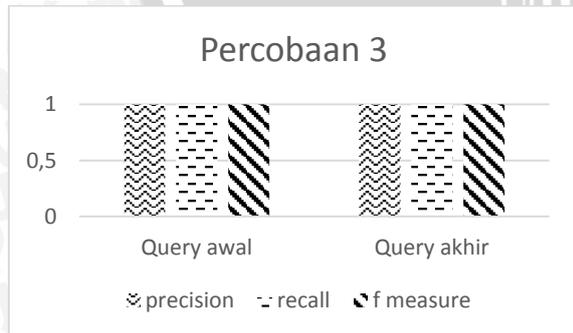
Grafik hasil *precision*, *recall* dan *f-measure* untuk percobaan 2.1 dengan *query* awal “sistem instalasi biodiesel” dan *query* akhir “sistem instalasi biodiesel bagus menarik” untuk metode TF.IDF.ICF bisa dilihat pada Gambar 8 dibawah ini :



Gambar 8 Grafik percobaan 2 TF.IDF.ICF

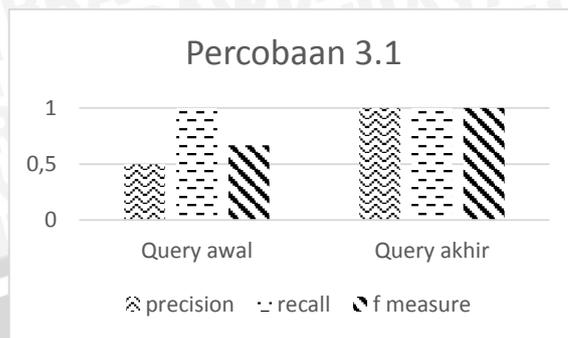
### 6.1.3 Hasil dan Analisa uji coba percobaan 3

Grafik hasil *precision*, *recall* dan *f-measure* untuk percobaan 3 dengan *query* awal “evaluasi ukuran panggung” dan *query* akhir “evaluasi ukuran panggung sempurna” untuk metode TF.IDF.ICSdF bisa dilihat pada Gambar 9 dibawah ini :



Gambar 9 Grafik percobaan 3 TF.IDF.ICSdF

Grafik hasil *precision*, *recall* dan *f-measure* untuk percobaan 3.1 dengan *query* awal “evaluasi ukuran panggung” dan *query* akhir “evaluasi ukuran panggung sempurna” untuk metode TF.IDF bisa dilihat pada Gambar 10 dibawah ini :



Gambar 10 Grafik percobaan 3.1 TF.IDF.ICF

Dari keseluruhan gambar grafik diatas dapat ditarik kesimpulan bahwa penambahan *query* dapat mempengaruhi hasil *precision*, *recall* dan *f-measure* baik untuk metode TF.IDF ataupun metode TF.IDF.ICSdF. Hasil *precision*, *recall* dan *f-measure* pada *query* akhir akan semakin bertambah.

### 6.3 Hasil dan Analisa uji coba pengujian 2

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.

## 7. KESIMPULAN DAN SARAN

### 7.1 Kesimpulan

Berdasarkan hasil uji coba parameter metode TF.IDF.ICSDf pada permasalahan perangkangan ide kreatif adalah sebagai berikut : Berdasarkan hasil uji coba parameter metode TF.IDF.ICSDf pada permasalahan perangkangan 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 perangkangan ide kreatif pada PT PJB UP Patiton 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 perankingan dokumen dengan menggunakan algoritma stemming yang terbaru. Disarankan tidak menggunakan kata yang mengandung *stoplist* yang tersimpan pada *filtering* maupun *stemming*.

## 8. DAFTAR PUSTAKA

- Fuji R., G.S. Mohammad, 2009. Class-indexing-based term weighting for automatic text classification, *Journal of Informetrics*.
- 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.
- Ledy Agusta. 2009. Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia, Fakultas Teknologi Informasi. Universitas Kristen Satya Wacana.
- Maulana, D. 2014. *Question Answering System Berbasis Clustering* pada Buku Pedoman PTIIK Dengan Algoritma Levenshtein Distance. Program Teknologi Informasi dan Ilmu Komputer.
- 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.pdf f ugm purwono [ diakses 21 September 2011].

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.ilic.nl/Publication/ResearchReport/Mol-200302.text.pdf](http://www.ilic.nl/Publication/ResearchReport/Mol-200302.text.pdf) [diakses 21 Februari 2016].

Taryana Suryana, 2012. Pengelolaan Dokumen Sebagai Sarana Komunikasi Internal Unikom. Program Studi Teknik Informatika. Universitas Komputer Indonesia