

APLIKASI WEBSITE DAN E-DMS KEJAKSAAN NEGERI KRAKSAAN DENGAN METODE TF-IDF

Akhmad Erieck Al Habsyi

Fakultas Teknik, Universitas Trunojoyo

Jl. Raya Telang PO. BOX 2 Kamal, Bangkalan, Madura, 691962

E-mail : eriksabella@gmail.com

ABSTRAK

Kejaksaan Negeri Kraksaan adalah lembaga yang melaksanakan kekuasaan negara, khususnya di bidang penuntutan perkara pidana maupun perdata. Dalam upaya mendukung kinerja Kejaksaan itu sendiri, diperlukan suatu sistem yang dapat mempublikasikan kejaksaan dan membantu dalam manajemen surat masuk maupun keluar, khususnya dalam pencarian suatu dokumen yang sesuai dengan query pengguna. Pada penelitian ini, penulis menggunakan metode Term Frequency Inverse Document Frequency (TF-IDF) yang menghasilkan pencarian dokumen yang sesuai tingkat pembobotan dan berdasarkan query masukan oleh pengguna. Metode Term Frequency Inverse Document Frequency (TF-IDF) didasarkan pada kemunculan term pada setiap dokumen dan pengurangan dominasi term yang banyak muncul di berbagai dokumen, sehingga mempermudah pengguna dalam melakukan pencarian suatu dokumen yang diinginkan secara cepat dan tepat. Hasil Penelitian ini adalah program yang manajemen surat dengan pencarian document teks yang menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF) yang menghasilkan perhitungan pembobotan Term Frequency Inverse Document Frequency (TF-IDF) dan mendapatkan dokumen relevan yang teranking sesuai tingkat pembobotannya berdasarkan query masukan oleh pengguna.

Kata Kunci : *Kejaksaan, Pembobotan , Query, TF-IDF, Pencarian Dokumen.*

ABSTRACT

Kraksaan state judiciary is the institution that accomplish the power state, particularly in the criminal and civil prosecution. In order to support the performance of those judiciary, it needs a system that can publish the judiciary and help to manage the mail either inbox or outbox, especially in searching of a document based on user query. In this study, the writer uses the method of Frequency Inverse Document Frequency (TF-IDF) that gets the appropriate document based on the level of weighting and user query. The method of Frequency Inverse Document Frequency (TF-IDF) is based on the term of each document and the term reducing in various documents, so it makes easier the user to search the document quickly and accurately. The result of this study is a program that can manage the mail in searching of the document by using the method of Frequency Inverse Document Frequency (TF-IDF) and produce the calculation of Term Frequency Inverse Document Frequency (TF-IDF), also gain the relevant document based on user query.

Keyword: *Judiciary, Weighting, Query, TF-IDF, Searching Document.*

PENDAHULUAN

Kejaksaan Republik Indonesia sebagai sebuah instansi Penegak Hukum yang berdiri dibawah naungan Kejaksaan Agung, yang memiliki visi yaitu memujudkan Kejaksaan sebagai lembaga penegak hukum yang melaksanakan tugasnya secara independen dengan menjunjung tinggi hak asasi manusia (HAM) dalam negara hukum berdasarkan Pancasila. Memiliki misi menyatukan tata pikir, tata laku dan tata kerja dalam penegakan hukum, optimalisasi pemberantasan KKN dan penuntasan pelanggaran HAM.

Banyak masyarakat yang kurang mengetahui hal tersebut, masyarakat kurang mengerti tentang fungsi dari kejaksaan itu sendiri, tugas kejaksaan disetiap sidang pengadilan, serta banyak hal yang menyangkut informasi tentang kejaksaan. Disamping itu, pada umumnya prosedur yang diterapkan oleh Kejaksaan Negeri dalam manajemen surat masuk dan surat keluar masih kurang efektif karena diterapkan secara manual. Dokumentasi surat masuk dan keluar hanya berupa penulisan di buku besar. Sedangkan pada tahap pengarsipan surat hanya berupa penyimpanan dokumen *hardcopy*.

Hal tersebut juga dialami oleh Kejaksaan Negeri Kraksaan yang membutuhkan media yang bisa menampung informasi mengenai Kejaksaan Kraksaan serta memanejemen surat masuk dan surat keluar secara tepat. Dengan bantuan website pada internet, informasi yang ingin disampaikan oleh instansi dapat tersampaikan kepada masyarakat luas. Serta dalam bidang memenajemen surat masuk dan surat keluar.

Untuk lebih efisiensi waktu dalam pencarian berkas perkara dari aplikasi ini, maka dalam pencarian berkas yang bisa menelusuri *query* pencarian sampai ke dalam isi berkas perkara yang di-*upload*.

Untuk menghasilkan berkas perkara yang relevan terhadap *query*

pencarian maka diperlukan proses perangkingan pada berkas perkara hasil pencarian. Metode yang digunakan untuk melakukan perangkingan adalah dengan menghitung bobot berkas terhadap *query*. Pembobotan ini menghitung bobot frekuensi kemunculan kata dalam suatu berkas (*term frequency – tf*) diimbangi dengan frekuensi kemunculan kata pada koleksi berkas (*inverse document frequency - idf*), yang dikenal dengan pembobotan *tf-idf*. Metode *Tf-idf* [1] merupakan suatu cara memberikan bobot hubungan suatu kata(*term*) terhadap berkas.

Oleh karena itu diperlukan adanya pengembangan sistem yang meliputi website dan sistem informasi manajemen surat masuk maupun keluar di Kejaksaan Negeri Kraksaan dengan menggunakan komputerisasi yang tentunya dapat meningkatkan kinerja instansi.

TINJAUAN PUSTAKA

1. Penelitian Sebelumnya
Izam Nurazwar Teknik Informaika ITS dalam penelitiannya yang berjudul “Pembangunan Aplikasi Document Management Sistem berbasis Web “, uji coba aplikasi dilakukan pada studi kasus manajemen dokumen penelitian mahasiswa informatika ITS dari hasil uji coba tersebut, Document Management System dapat memudahkan pembuatan, klasifikasi, pengaturan hak akses dan pencarian dokumen[2]. Pada penelitian ini, penulis mengkombinasikan pembuatan document management system dengan pencarian menggunakan metode TF-IDF dan Website Kejaksaan Negeri Kraksaan.

Irwin Darmawan, Teknik Informatika, Universitas Trunojoyo, dalam penelitiannya “Rancang Bangun Aplikasi Keyword Generator Untuk Mendukung Pencarian Documen Menggunakan TF-IDF”, dibangun suatu sistem keyword generator berbasis java yang mampu mengekstraksi Text dari 4 jenis dokumen yang berekstensi *.pdf, *.txt, *.docx dan *.doc. Adapun

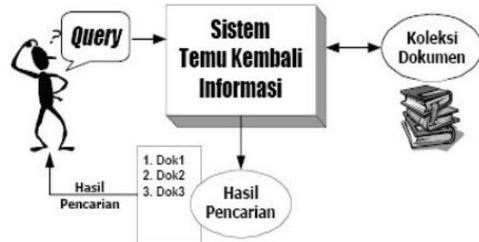
pencarian document dilakukan melalui aplikasi berbasis web dengan menggunakan metode TF-IDF[3]. Didalam penelitian ini, penulis mengimplementasikan pencarian dokumen yang dapat diekstraksi dan menggunakan metode TF-IDF sebagai acuan dasar untuk proses pembobotan term sebagai ukuran kemiripan teks dalam proses pencarian.

2. E-DMS (Elektronic-Document Management System)

Sebuah Sistem Manajemen Dokumen adalah sistem yang berfungsi untuk memfasilitasi penyimpanan dan ekstraksi dokumen dari database yang berhubungan dengan proses menghasilkan kinerja yang efisiensi dan maksimal[4]. Sebuah Sistem Manajemen Dokumen Elektronik (EDMS) menyimpan versi elektronik dari semua dokumen Anda, salinan biasanya scan dokumen kertas dan dokumen

3. Sistem Temu Baik Informasi

Temu kembali informasi adalah adalah ilmu pencarian informasi pada dokumen, pencarian untuk dokumen itu sendiri, pencarian untuk metadata yang menjelaskan dokumen, atau mencari di dalam *database*, baik relasi *database* yang *stand-alone* atau *hiperText database* yang terdapat pada *network* seperti internet atau *World Wide Web* atau intranet, untuk teks, suara, gambar, atau data[5].



Gambar 1. Sistem Temu Kembali Informasi

4. Text Mining

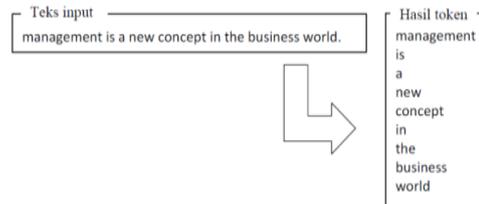
Text mining adalah suatu proses menggali informasi dari data yang berupa teks dimana sumber data biasanya didapatkan dari dokumen[6].

5. Proses Penentuan Indeks

Langkah-langkah yang dilakukan dalam proses penentuan index adalah

a. *Tokenizing*

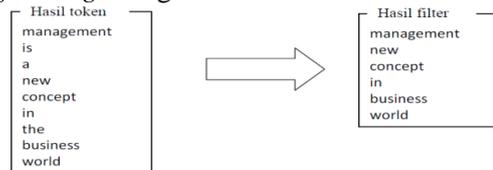
Merupakan proses memecah (*parsing*) file menjadi *token-token* yaitu dengan memotong menjadi *term*[7]. Contoh hasil *tokenizing* sebagai berikut :



Gambar 2. Hasil *Tokenizing*[7]

b. *Filtering*

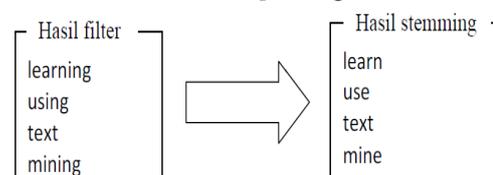
Merupakan proses menghilangkan kata-kata umum yang sering ditampilkan dalam *file* seperti : *is*, *for*, *in*, dan sebagainya[7]. Contoh hasil *filtering* sebagai berikut:



Gambar 3. Hasil *Filtering* [7]

c. *Stemming*

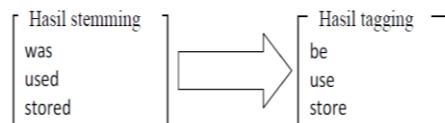
Merupakan proses pembuangan *prefix* (awalan) dan *suffix* (imbuan akhiran) secara morfologi dari suatu kata berimbuan menjadi kata dasar[7]. Contoh hasil *stemming* sebagai berikut :



Gambar 4. Hasil *Stemming* [7]

d. *Tagging*

Tahapan *tagging* adalah tahap mencari bentuk awal/root kata dari tiap kata dalam bentuk lampau atau kata hasil *stemming*[7]. Hasil *tagging* sebagai berikut:



Gambar 5. Hasil *Tagging* [7]

e. *Term weighting*

Merupakan proses pembobotan *term* yang ada di *file* (seberapa sering kemunculan *term* pada *file*)[7]. Contoh proses penentuan indeks :

1. Teks yang diproses adalah "Text mining, sometimes alternately referred to as text data mining"
2. Proses *tokenizing* terhadap teks, maka dihasilkan kata kata sebagai berikut :

```
text
mining
sometimes
alternately
referred
to
as
text
data
mining
```

Gambar 6. Hasil *Tokenizing* [7]

3. Setelah proses *tokenizing*, dilakukan proses *filtering* yang berfungsi untuk menghilangkan kata-kata umum. Dalam kasus ini yang menjadi kata umum adalah "as" dan "to". Daftar kata-kata umum bisa dilihat pada lampiran. Hasil *filtering* sebagai berikut :

```
text
mining
sometimes
alternately
referred
text
data
mining
```

Gambar 7. Hasil *Filtering* [7]

4. Proses *stemming*. Sehingga dihasilkan hasil proses dari *stemming* sebagai berikut:

```
text
mine
sometime
altern
refer
text
data
mine
```

Gambar 8. Hasil *stemming* [7]

5. Selain dilakukan proses *stemming*, untuk kata yang tidak ada dalam kamus maka dilakukan proses *tagging*. Sehingga dihasilkan daftar sebagai berikut:

```
text
mine
sometime
altern
refer
text
data
mine
```

Gambar 9. Hasil *Tagging* [7]

6. Proses terakhir adalah menghitung *term weighting* dari kata-kata yang telah diproses. Hasil perhitungannya sebagai berikut:

Kata	Frekuensi
text	2
mine	2
sometime	1
altern	1
refer	1
data	1

Gambar 10. Hasil *Tagging* [7]

6. Proses Pencarian

Tahapan dalam proses pencarian adalah:

a. *Tokenizing*

Tahap awal dalam melakukan pencarian adalah memecah (*tokenizing*) *query*, *keyword* yang dimasukkan pengguna dijadikan sebagai *query* untuk melakukan pencarian kedalam *database*.

b. *Filtering*

Proses yang dilakukan untuk menghilangkan kata-kata yang umum.

c. *Stemming*

Setelah *query* tersebut melewati proses *filtering*, kemudian *query* tersebut melalui proses *stemming* yaitu suatu proses membuang awalan, akhiran dari kata.

d. *Tagging*

Proses mencari bentuk awal atau root kata dari tiap kata dalam bentuk lampau atau kata hasil *stemming*.

e. *Ranking*

Tahap yang terakhir adalah melakukan perangkingan, dari *file-file* yang didapatkan. Perangkingan disusun berdasarkan bobot masing-masing *file* terhadap kata kunci, *file* yang paling besar bobotnya menjadi *file* dengan *ranking* teratas.

7. **TF-IDF**

Metode Tf-idf [1] merupakan suatu cara memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata didalam sebuah dokumen tertentu dan inverse frekuensi dokumen yang mengandung kata tersebut.

Rumus Umum Tf-Idf

$$W_{ij} = tf_{ij} \cdot idf_i$$

$$W_{ij} = tf_{ij} \log_2 (N/n)$$

Keterangan :

w_{ij} = bobot kata/term t_j terhadap dokumen d_i

tf_{ij} = jumlah kata/term t_j dalam d_i

N = jumlah dokumen yang ada

n = jumlah dokumen yang mengandung kata/term t_j

(minimal ada satu kata yaitu term t_j)

Berdasarkan rumus diatas, berapapun besarnya nilai tf_{ij} , apabila $N = n$ maka akan didapatkan hasil 0 (nol) untuk perhitungan Idf. Untuk itu dapat ditambahkan nilai 1 pada sisi Idf, sehingga perhitungan bobotnya menjadi : $W_{ij} = tf_{ij} (\log (N/n) + 1)$

8. Pengertian Kejaksaan

Kejaksaan RI adalah lembaga yang melaksanakan kekuasaan negara, khususnya di bidang penuntutan. Sebagai badan yang berwenang dalam penegakan hukum dan keadilan, Kejaksaan dipimpin oleh Jaksa Agung yang dipilih oleh dan bertanggung jawab kepada Presiden. Kejaksaan Agung, Kejaksaan Tinggi, dan Kejaksaan Negeri merupakan kekuasaan negara khususnya dibidang penuntutan, dimana semuanya merupakan satu kesatuan yang utuh dan tidak dapat dipisahkan[8].

METODE

Metodologi penelitian yang digunakan dalam penelitian ini adalah :

1. Metode Observasi

Metode ini diterapkan dengan cara melakukan pengamatan langsung pada Kejaksaan Negeri Kraksaan guna mendapatkan gambaran jelas dan terkait terhadap permasalahan sehingga dapat diketahui kebutuhan yang diperlukan.

2. Metode Interview

Proses tanya jawab mengenai permasalahan dimana dua orang atau lebih bertatap muka mendengarkan secara langsung informasi atau keterangan sehingga penelitian yang dibuat tepat sasaran.

3. Metode Analisa

Menganalisa masalah masalah yang akan disajikan dan mengumpulkan data atau informasi.

4. Metode Literature

Usaha untuk lebih memudahkan dalam melengkapi data memecahkan masalah.

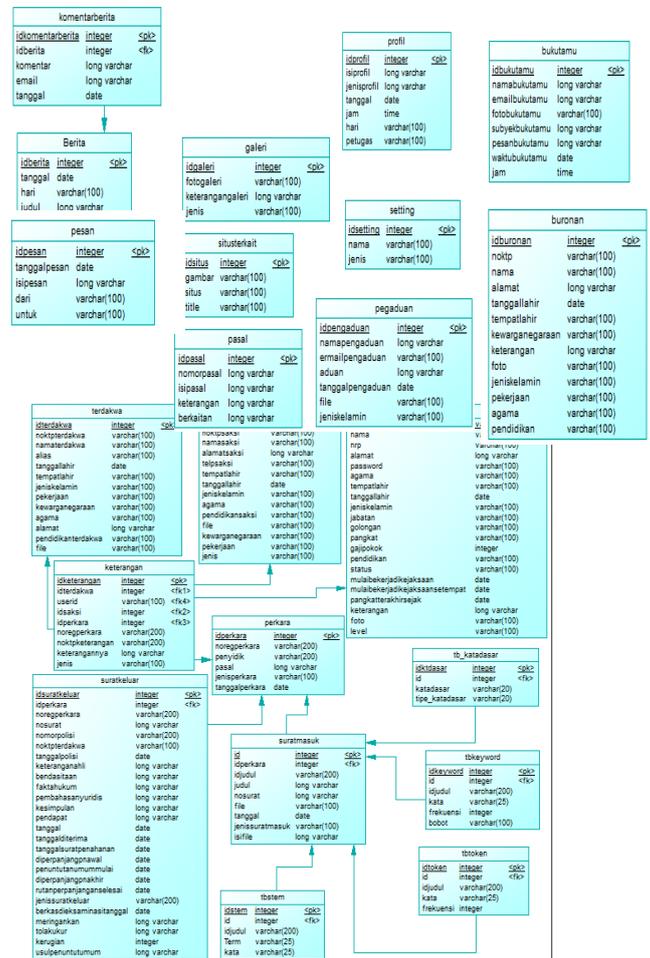
5. Metode Evaluasi

Mengevaluasi hasil-hasil yang telah dikerjakan.

HASIL DAN PEMBAHASAN

1. PDM (Physical Data Model)

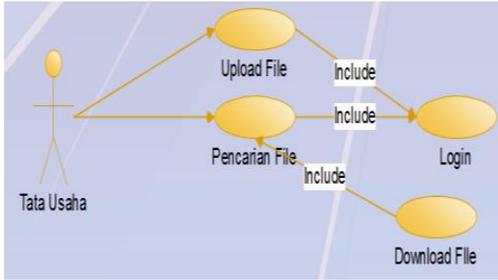
Model yang menggunakan sejumlah tabel untuk menggambarkan data serta hubungan antara data-data tersebut.



Gambar 11. Physical Data Model.

2. Use Case Diagram Aplikasi E-DMS

Diagram ini menggambarkan fungsional dari aplikasi E-DMS.



Gambar 12. Use Case E-DMS.

3. Uji Coba

Dalam uji coba sistem ini digunakan data uji coba sebagai berikut :

Contoh simulasi perhitungan nilai $tf \cdot idf$ bisa dilihat pada bagian dibawah ini :

Tabel 1. Contoh File, Isifile dan Hasil Penentuan Index

File	Text	Terms
File1.pdf	kepala desa sekar kare telah melakukan tindak pidana korupsi alokasi dana desa.	kepala desa sekar kare tindak pidana korupsi alokasi dana
File2.doc	polisi telah melakukan penyidikan terhadap tersangka tindak pidana korupsi alokasi dana desa yang dilakukan tersangka.	polisi penyidik tersangka tindak pidana korupsi alokasi dana desa
File3.docx	polisi membawa tersangka tindak pidana korupsi atas nama bukamin ke kantor polisi probolinggo	polisi tersangka tindak pidana korupsi nama bukamin kantor probolinggo
File4.txt	sebagai kepala desa, bukamin telah menggunakan dana alokasi desa untuk perhal pribadi	kepala desa bukamin dana alokasi perhal pribadi

Tabel 2. Tabel Penentuan Jumlah Kata Setelah Penentuan Index Hitung document frequency (df)

	kepala	desa	sekar	kare	tindak	pidana	korupsi
File1.pdf	1	2	1	1	1	1	1
File2.doc		1			1	1	1
File3.docx					1	1	1
File4.txt	1	1					

	alokasi	dana	polisi	penyidik	Tersangka	nama
File1.pdf	1	1				
File2.doc	1	1	1	1	2	
File3.docx			1		1	1
File4.txt	1	2				

	bukamin	kantor	probolinggo	perhal	pnbadi
File1.pdf					
File2.doc					
File3.docx	1	1	1		
File4.txt	1			1	1

Tabel 3. Tabel Frekuensi dari Term

Term	document frequency (df)
Kepala	2
Desa	3
Sekar	1
Kare	1
Tindak	3
Pidana	3
Korupsi	3
Alokasi	3
Dana	3
Polisi	2
Penyidik	1
Tersangka	2
Nama	1
Bukamin	2
Kantor	1
Probolinggo	1
Perihal	1
Pribadi	1

Hitung *invers document frequency (idf)*

Tabel 4. Tabel Penentuan Nilai Idf Untuk Setiap Term dari 4 Dokumen.

Term	document frequency (df)	Invers document frequency (idf)
Kepala	2	$\log(4/2)+1=1,30$
Desa	3	$\log(4/3)+1=1,13$
Sekar	1	$\log(4/1)+1=1,60$
Kare	1	$\log(4/1)+1=1,60$
Tindak	3	$\log(4/3)+1=1,13$
Pidana	3	$\log(4/3)+1=1,13$
Korupsi	3	$\log(4/3)+1=1,13$
Alokasi	3	$\log(4/3)+1=1,13$
Dana	3	$\log(4/3)+1=1,13$
Polisi	2	$\log(4/2)+1=1,30$
Penyidik	1	$\log(4/1)+1=1,60$
Tersangka	2	$\log(4/2)+1=1,30$
Nama	1	$\log(4/1)+1=1,60$
Bukamin	2	$\log(4/2)+1=1,30$
Kantor	1	$\log(4/1)+1=1,60$
Probolinggo	1	$\log(4/1)+1=1,60$
Perihal	1	$\log(4/1)+1=1,60$
Pribadi	1	$\log(4/1)+1=1,60$

Hitung *document frequency (df)-invers document frequency (idf)*

Tabel 5. Tabel Penentuan Bobot untuk setiap Term dari 4 Dokumen tersebut

term	idf	$\frac{df}{DF_{File1.pdf}}$	$\frac{df}{DF_{File2.doc}}$	$\frac{df}{DF_{File3.docx}}$	$\frac{df}{DF_{File4.txt}}$	$\frac{idf}{DF_{File1.pdf}}$	$\frac{idf}{DF_{File2.doc}}$	$\frac{idf}{DF_{File3.docx}}$	$\frac{idf}{DF_{File4.txt}}$
kepala	$\log(4/2)+1=1,30$	1				1,30			1,30
desa	$\log(4/3)+1=1,13$	2	1			2,26	1,13		1,13
sekar	$\log(4/1)+1=1,60$	1				1,60			
kare	$\log(4/1)+1=1,60$	1				1,60			
tindak	$\log(4/3)+1=1,13$	1	1	1		1,13	1,13	1,13	
pidana	$\log(4/3)+1=1,13$	1	1	1		1,13	1,13	1,13	
korupsi	$\log(4/3)+1=1,13$	1	1	1		1,13	1,13	1,13	
alokasi	$\log(4/3)+1=1,13$	1	1		1	1,13	1,13		1,13
dana	$\log(4/3)+1=1,13$	1	1		2	1,13	1,13		2,26
polisi	$\log(4/2)+1=1,30$		1	1			1,30	1,30	
penyidik	$\log(4/1)+1=1,60$		1				1,60		
tersangka	$\log(4/2)+1=1,30$		2	1			2,60	1,30	
nama	$\log(4/1)+1=1,60$			1				1,60	
bukamin	$\log(4/2)+1=1,30$			1	1			1,30	1,30
kantor	$\log(4/1)+1=1,60$				1				1,60
probolinggo	$\log(4/1)+1=1,60$				1				1,60
perihal	$\log(4/1)+1=1,60$					1			1,60
pribadi	$\log(4/1)+1=1,60$					1			1,60
$\sum idf$						12,41	12,28	12,48	10,32

Maka, nilai $tf \cdot idf$ berdasarkan query diatas adalah :

File1.pdf : 12,41
File2.doc : 12,28
File3.docx : 12,48
File4.txt : 10,32

SIMPULAN

Beberapa kesimpulan yang dapat diambil dari perancangan dan pembuatan aplikasi ini adalah :

- a. Aplikasi ini dapat mempermudah proses pembuatan surat dan manajemen surat perkara pidana khusus maupun pidana umum.
- b. Mempermudah mempublikasikan informasi tentang kejahatan terhadap masyarakat khususnya masyarakat Kraksaan, Probolinggo.
- c. Aplikasi ini dapat membantu pengguna untuk dapat melakukan proses pencarian dokumen dengan menggunakan metode tf-idf. Metode tf-idf merupakan proses menghitung frekuensi kemunculan *term* dalam sebuah *file*, dibandingkan dengan statistik kemunculan term pada keseluruhan koleksi *file*.

SARAN

Berikut ini merupakan saran untuk mengembangkan proyek penelitian ini untuk selanjutnya :

- a. Aplikasi ini dapat dikembangkan lebih lanjut dengan menambahkan fasilitas-fasilitas lain guna mendukung aplikasi ini di kemudian hari.
- b. Sarana penunjang aplikasi yang meliputi perangkat komputer beserta jaringannya terlebih dahulu harus dipersiapkan agar penerapan sistem dapat lebih optimal.

DAFTAR PUSTAKA

- [1] Intan, R, Defeng A. 2006. *HARD Subject-base Search Eagler menggunakan TF-IDF dan Jaccard's*. Surabaya. Universitas Petra.
- [2] Nurazwar, Izam. 2007. *Pembangunan Aplikasi Document*

Management System Berbasis Web. Tugas Akhir Surabaya. ITS

- [3] Darmawan, Irwan. 2010. *Rancang Bangun Aplikasi Keyword Generator Untuk Mendukung Pencarian Dokumen Menggunakan Metode TF-IDF*. Tugas Akhir Bangkalan : Universitas Trunojoyo.
- [4] Jusoff, Kamaruzaman dkk. 2011. *Electronic Document Management System*. World Applied Sciences Journal 12.
- [5] Chantique, Dyah. 2007. *Pengantar Temu Kembali Informasi Retrieval*. <http://blog.its.ac.id/dyah03tc/2007/11/21/pengantar-temu-kembali-informasi-information-retrieval/>. Diakses tanggal 02 Februari 2013.
- [6] Khodra, Masayu Leylia. 2003. *Text Mining Kategorisasi Teks Naïve Bayes*. Bandung: IF-ITB.