



**PENERAPAN METODE TF-IDF DAN COSINE
SIMILARITY DALAM APLIKASI KITAB UNDANG-
UNDANG HUKUM DAGANG**

Skripsi

**diajukan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana
Pendidikan Program Studi Pendidikan Teknik Informatika dan Komputer**

Oleh

Hirroe Wijaya Ani Kesuma NIM. 5302411174

UNNES
UNIVERSITAS NEGERI SEMARANG
JURUSAN TEKNIK ELEKTRO

FAKULTAS TEKNIK

UNIVERSITAS NEGERI SEMARANG

2016



LEMBAR KEASLIAN

Dengan ini saya menyatakan bahwa :

1. Skripsi ini, adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana, magister, dan/atau doktor), baik di Universitas Negeri Semarang (UNNES) maupun di perguruan tinggi lain.
2. Karya tulis ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan Pembimbing dan masukan Tim Penguji.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya ini, serta sanksi lainnya sesuai dengan norma yang berlaku di perguruan tinggi ini.



Semarang, 14 Desember 2015

Hirroe Wijaya Ani Kesuma
NIM.5302411174

LEMBAR PERSETUJUAN PEMBIMBING

Nama : Hirroe Wijaya Ani Kesuma

NIM : 5302411174

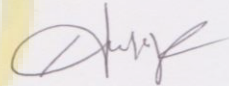
Program Studi : S-1 Pendidikan Teknik Informatika dan Komputer

Judul Skripsi : Penerapan Metode TF-IDF dan Cosine Similarity dalam Aplikasi Kitab Undang-Undang Hukum Dagang

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang panitia ujian skripsi Program Studi S-1 Pendidikan Teknik Informatika dan Komputer FT. UNNES.

Semarang, 14 Desember 2015

Pembimbing,



Feddy Setio Pribadi

NIP. 197808222003121002

UNNES
UNIVERSITAS NEGERI SEMARANG

LEMBAR PENGESAHAN

Skripsi dengan Judul “Penerapan Metode TF-IDF dan Cosine Similarity dalam Aplikasi Kitab Undang-Undang Hukum Dagang” telah dipertahankan di depan sidang Panitia Ujian Skripsi Fakultas Teknik UNNES pada tanggal 25 Januari 2016.

Oleh :

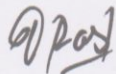
Nama : Hirroe Wijaya Ani Kesuma

NIM : 5302411174

Program Studi : Pendidikan Teknik Informatika dan Komputer

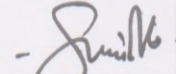
Panitia :

Ketua Panitia



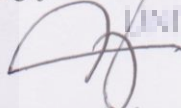
Dr.-Ing. Dhidik Prastiyanto, S.T., M.T.
NIP. 19780531 200501 1 002

Sekretaris



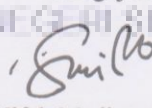
Ir. Ulfah Mediaty Arief, M.T.
NIP. 19660505 199802 2 001

Penguji I



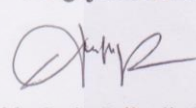
Tatyantoro Andrasto, S.T., M.T.
NIP. 19680316 199903 1 001

Penguji II



Ir. Ulfah Mediaty Arief, M.T.
NIP. 19660505 199802 2 001

Penguji III/Pembimbing



Feddy Setio Pribadi, S.Pd., M.T.
NIP. 19780822 200312 1 002

Mengetahui,

Dekan Fakultas Teknik,



Dr. Nur Qudus, M.T.

NIP. 19691130 199403 1 001

MOTTO DAN PERSEMBAHAN

1. *“Karena sesungguhnya sesudah kesulitan itu ada kemudahan. Sesungguhnya sesudah kesulitan itu ada kemudahan.” (Q. S. Al-Insyirah ayat 5-6)*
2. *“Don’t use words too big for the subject. Don’t say infinitely when you mean very, otherwise you’ll have no word left when you want to talk about something really infinite.” (C. S. Lewis)*

Dengan mengucap syukur Alhamdulillah, kupersembahkan karya kecilku ini untuk orang-orang yang kusayangi :

- Ibu Bapak tercinta, yang tak pernah lelah memberikan doa, pengorbanan, dukungan, dan kesabarannya hingga mengantarkanku sampai saat ini.
- Budhe dan Pakdhe yang sudah mendukung dalam banyak hal.
- Adikku Lia, Oos, Dedek dan keponakan tersayang yang selalu memberikan semangat dan doa.
- ION-ku tersayang yang banyak memberikan dukungan.
- Sahabat-sahabatku seperjuangan dan teman-teman PTIK 2011 yang selalu membantu. Terimakasih atas waktu dan dukungannya.

ABSTRAK

Kesuma, Hirroe Wijaya Ani. 2015. “Penerapan Metode TF-IDF dan Cosine Similarity dalam Aplikasi Kitab Undang-Undang Hukum Dagang”.

Pembimbing : Feddy Setio Pribadi, S.Pd., M.T. Prodi : Pendidikan Teknik Informatika dan Komputer.

Kitab Undang-Undang Hukum Dagang (KUHD) merupakan suatu kitab yang mengatur tentang hukum dagang di Indonesia. Selama ini cara untuk mencari suatu pasal berdasarkan suatu kata kunci hanya beracuan pada kata yang sudah diindekskan di dalam buku, seperti yang terjadi pada Aplikasi Kitab Undang-Undang Hukum Pidana (KUHP) versi Android. Penelitian ini bertujuan untuk membangun sistem temu balik informasi (*Information Retrieval*) pada Kitab Undang-Undang Hukum Dagang yang diharapkan fungsinya lebih baik dari sebelumnya.

Alur dari aplikasi ini dimulai dengan tahap *preprocessing* terhadap dokumen dan *query*. Proses dilanjutkan dengan pembobotan kata menggunakan metode TF-IDF. Kemudian dilakukan proses pencocokan dan perhitungan kedekatan antara *query* dengan dokumen pasal menggunakan metode *Cosine similarity*. Penelitian dilanjutkan dengan melakukan pengujian diantaranya pengujian *Blackbox* (untuk menguji sistem), uji pakar, *performance measure* (untuk menguji kinerja sistem), dan uji korelasi (untuk menguji relevansi hasil).

Hasil dari pengujian *Blackbox* menunjukkan bahwa jumlah kata maksimal yang bisa diinput adalah 4 buah kombinasi kata, dan sistem tidak dapat melakukan pengindeksan kata seperti pada kitab. Sementara itu pengujian pakar dan *performance measure* menghasilkan 55,04% untuk *recall*, 63,33% untuk *precision*, dan 56,93% untuk *f-measure*. Untuk besarnya hasil uji korelasi adalah 0,71. Secara umum metode TF-IDF dan *Cosine Similarity* cukup baik untuk diterapkan dalam aplikasi.

Kata Kunci : *Information Retrieval, TF-IDF, Cosine Similarity, pengujian Blackbox, uji pakar, performance measure, Kitab Undang-Undang Hukum Dagang*

KATA PENGANTAR

Puji syukur Alhamdulillah penulis panjatkan kehadiran Allah SWT yang telah melimpahkan nikmat, rahmat, taufik, dan hidayah-Nya, sehingga penulisan skripsi ini dapat diselesaikan dengan baik.

Laporan skripsi ini disusun guna memenuhi salah satu syarat untuk menyelesaikan Program Studi Pendidikan Teknik Informatika dan Komputer Jurusan Teknik Elektro Fakultas Teknik Universitas Negeri Semarang.

Keberhasilan penulis dalam menyusun skripsi ini tidak terlepas dari bantuan, dorongan, serta doa dari berbagai pihak baik secara langsung maupun tidak langsung. Pada kesempatan ini, dengan segala kerendahan hati penulis menyampaikan terima kasih yang setulus-tulusnya kepada :

1. Prof. Dr. Fathur Rokhman, M.Hum., sebagai Rektor Universitas Negeri Semarang.
2. Dr. Nur Qudus, M.T., sebagai Dekan Fakultas Teknik Universitas Negeri Semarang.
3. Dr.-Ing. Dhidik Prastiyanto, S.T., M.T., sebagai Ketua Jurusan Teknik Elektro.
4. Ir. Ulfah Mediaty Arief, M.T., sebagai Ketua Program Studi Pendidikan Teknik Informatika dan Komputer.
5. Feddy Setio Pribadi, S.Pd., M.T., sebagai dosen pembimbing yang telah memberikan arahan, bimbingan, dan saran kepada penulis dengan teliti dan sabar.

6. Seluruh Dosen dan Staf Karyawan Jurusan Teknik Elektro.
7. Nurul Fibrianti, S.H., M.Hum., sebagai sumber informasi dan penguji pakar aplikasi yang sudah banyak memberikan arahan untuk keberhasilan penelitian ini.
8. Keluarga tercinta, Ibu, Bapak dan Adik-adik yang selalu memberikan doa, semangat, serta dorongan yang tiada hentinya.
9. Delia, Hendro, Mida, Alfian, Ina, Fitri dan semua rekan PTIK 2011 serta teman-teman Kos Priyangan dan Kos V2 yang selalu memberikan bantuan serta semangat dalam penyusunan skripsi ini.

Semoga laporan skripsi ini dapat bermanfaat bagi semua pihak sebagaimana yang diharapkan. Aamin.

Penulis,

Hirroe Wijaya Ani Kesuma



DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR KEASLIAN	iii
LEMBAR PERSETUJUAN PEMBIMBING	iv
LEMBAR PENGESAHAN	v
MOTTO DAN PERSEMBAHAN	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMPIRAN	xvii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	5
1.4 Batasan Masalah	5
1.5 Tujuan Penelitian	5

1.6	Manfaat Penelitian.....	6
1.7	Penegasan Istilah	7
BAB II TINJAUAN PUSTAKA.....		8
2.1	Landasan Teori	8
2.1.1	Kitab Undang-Undang Hukum Dagang	8
2.1.2	Aplikasi.....	9
2.1.3	Konsep Dasar Sistem Temu Balik Informasi	9
2.1.4	Metode TF-IDF (<i>Term Frequency – Invers Document Frequency</i>).....	13
2.1.5	Cosine Similarity	14
2.1.6	Evaluasi Sistem Information Retrieval	15
2.1.7	Sistem Temu Balik Informasi “Kitab Undang-Undang Hukum Dagang” Menggunakan Metode TF-IDF dan Cosine Similarity	19
2.1.8	Penelitian Terdahulu	21
2.1.9	Perangkat Pembangun Aplikasi	24
2.2	Kerangka Berfikir.....	25
BAB III METODOLOGI PENELITIAN		28
3.1	Metode Pengumpulan Data	28
3.1.1	Observasi	28
3.1.2	Studi Pustaka.....	34

3.2 Konsep Dasar Sistem Temu Balik Informasi	34
3.2.1 Arsitektur Sistem	34
3.2.2 Desain	35
3.3 Perancangan Pengujian.....	38
3.3.1 Bahan Pengujian	38
3.3.2 Tujuan Pengujian	39
3.3.3 Skenario dan Kriteria Pengujian	39
BAB IV HASIL DAN PEMBAHASAN	41
4.1 Hasil Pengujian.....	41
4.1.1 <i>Documents Collection</i>	41
4.1.2 <i>Preprocessing</i>	41
4.1.3 <i>TF-IDF (Term Frequency – Inverse Document Frequency)</i>	42
4.1.4 <i>Cosine Similarity</i>	44
4.1.5 <i>Pengujian Blackbox</i>	47
4.1.6 <i>Expert Judgement (Uji Pakar)</i>	54
4.1.7 <i>Performance Measure</i>	55
4.1.8 Uji Korelasi Pearson	58
4.2 Pembahasan	59
BAB V PENUTUP.....	70
5.1 Kesimpulan.....	70

5.2	Saran	70
DAFTAR PUSTAKA		72
LAMPIRAN.....		76



DAFTAR TABEL

	Halaman
Tabel 4.1 Pengujian <i>Blackbox</i>	48
Tabel 4.2 Tabel Hasil Uji Pakar untuk <i>Performance Measure</i>	54
Tabel 4.3 Tabel Hasil Uji Pakar untuk Korelasi Pearson.....	55
Tabel 4.4 Nilai <i>Recall</i>	56
Tabel 4.5 Nilai <i>Precision</i>	57
Tabel 4.6 Nilai <i>F-measure</i>	58
Tabel 4.7 Tabel Perhitungan untuk Uji Korelasi Pearson.....	59
Tabel 4.8 Perbandingan hasil output pada “waris” dan “ahli waris”	63
Tabel 4.9 Hasil perhitungan pencocokan untuk kata “waris”	64
Tabel 4.10 Hasil perhitungan pencocokan untuk frase “ahli waris”	65



DAFTAR GAMBAR

	Halaman
Gambar 2.1 Arsitektur Sistem Temu Balik Informasi	10
Gambar 2.2 Proses <i>Tokenization</i>	12
Gambar 2.3 Tahap <i>tokenizing</i> pasal	20
Gambar 2.4 Tahap <i>filtering</i> pasal.....	20
Gambar 2.5 Kerangka Berfikir.....	27
Gambar 3.1 Tampilan halaman awal dari Aplikasi Kitab Undang-Undang Hukum Pidana.....	29
Gambar 3.2 Tampilan kolom pencarian.....	30
Gambar 3.3 Tampilan hasil dari pencarian menggunakan keyword “pembunuhan”	30
Gambar 3.4 Tampilan hasil dari pencarian menggunakan keyword “pencurian dan pembunuhan”	31
Gambar 3.5 Kata “Komunisme/Marxisme” muncul secara berurutan.....	32
Gambar 3.6 Tidak terdapat hasil untuk keyword “komunisme marxisme”	33
Gambar 3.7 Arsitektur Aplikasi Kitab Undang-Undang Hukum Dagang	35
Gambar 3.8 Alur Proses Aplikasi Kitab Undang-Undang Hukum Dagang	36
Gambar 3.9 <i>Interface</i> Halaman Awal	37
Gambar 3.10 <i>Interface</i> Hasil Pencarian	38
Gambar 4.1 Hasil <i>preprocessing</i> sistem.....	42
Gambar 4.2 Tabel Perhitungan TF-IDF	43

Gambar 4.3 Hasil pencarian untuk <i>query</i> “kasir”	44
Gambar 4.4 Hasil pencarian untuk <i>query</i> “perseroan terbatas”	45
Gambar 4.5 Hasil pencarian <i>query</i> “hak pimpinan firma”	46
Gambar 4.6 Hasil pencarian untuk <i>query</i> “hak waris perusahaan dagang”	47



DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Daftar Kata <i>Stopword</i>	77
Lampiran 2 Hasil Uji Pakar untuk <i>Performance Measure</i> Aplikasi Kitab Undang-Undang Hukum Dagang	81
Lampiran 3 Hasil Uji Pakar untuk Korelasi Pearson Aplikasi Kitab Undang-Undang Hukum Dagang	85
Lampiran 4 Hasil Uji <i>Blackbox</i>	88
Lampiran 5 Form Usul Topik	97
Lampiran 6 Surat Usulan Pembimbing	98
Lampiran 7 Surat Keputusan Dosen Pembimbing	99
Lampiran 8 Surat Tugas Penguji	100



BAB I

PENDAHULUAN

1.1 Latar Belakang

Hukum Dagang adalah hukum yang mengatur tingkah laku manusia yang turut melakukan perdagangan dalam usahanya memperoleh keuntungan. Dapat juga dikatakan, Hukum Dagang ialah hukum yang mengatur hubungan hukum antara manusia-manusia dan badan-badan hukum satu sama lainnya, dalam lapangan perdagangan (Kansil, 2008:20).

Kitab Undang-Undang Hukum Dagang (KUHD) Indonesia kira-kira satu abad yang lalu telah dibawa orang Belanda ke tanah air kita, mula-mula ia hanya berlaku bagi orang-orang Eropa di Indonesia (berdasarkan asas konkordansi). Kemudian juga dinyatakan berlaku bagi orang-orang Timur Asing, akan tetapi tidak berlaku seluruhnya untuk orang-orang Indonesia (hanya bagian-bagian tertentu saja).

KUHD yang mulai berlaku di Indonesia pada 1 Mei 1848 terbagi atas dua kitab dan 23 bab. Kitab I terdiri dari 10 bab dan Kitab II terdiri dari 13 bab (Kansil, 2008:21).

Permasalahan yang muncul adalah mencari bagian pasal maupun ayat yang dilakukan secara manual. Dengan kemajuan teknologi yang sangat pesat pada saat ini dibutuhkan sebuah aplikasi yang dapat mengakomodir kebutuhan setiap pengguna akan sebuah aplikasi pengganti buku yang mudah dibawa serta dapat digunakan kapanpun dan dimanapun dengan efektif (Harnanto, 2013).

Aplikasi adalah *software* yang dibeli perusahaan dari tempat pembuatnya. Aplikasi akan menggunakan sistem operasi (SO) komputer dan aplikasi lainnya yang mendukung. Istilah ini mulai perlahan masuk dalam istilah Teknologi Informasi semenjak tahun 1993, yang biasanya juga disingkat dengan app. Secara historis, aplikasi adalah *software* yang dikembangkan oleh sebuah perusahaan. *Software* aplikasi adalah program yang menentukan aktifitas pemrosesan informasi yang dibutuhkan untuk penyelesaian tugas-tugas khusus dari pemakai komputer (Indrajani, 2009:5).

Sebelumnya sudah ada aplikasi Kitab Undang-Undang Hukum Pidana (KUHP) yang dapat dijalankan pada sistem operasi *Android* yang memuat tentang aturan pidana. Secara umum, struktur pengelompokan pasal dalam Kitab Undang-Undang Hukum Pidana dan Kitab Undang-Undang Hukum Dagang memiliki kesamaan. Pasal yang ada dalam Kitab Undang-Undang Hukum Pidana juga dikelompokkan dalam beberapa bab dan beberapa buku. Terdapat tiga buku dan sejumlah bab yang disusun sesuai indeks yang telah ditetapkan.

Cara kerja Aplikasi Kitab Undang-Undang Hukum Pidana adalah dengan memasukkan satu *keyword* yang diinginkan untuk dicocokkan dengan pasal-pasal yang ada dalam Kitab Undang-Undang Hukum Pidana. Fungsi kerja yang demikian dikenal dengan istilah *information retrieval* atau sistem pencarian informasi. Namun, ketika dimasukkan dua atau lebih *keyword*, Aplikasi Kitab Undang-Undang Hukum Pidana ini tidak memberikan hasil yang diinginkan, meskipun kombinasi kata kunci merupakan kata-kata yang masih dalam satu

pasal. Sehingga masih perlu dilakukan pengembangan terhadap aplikasi yang nantinya akan diterapkan pada Kitab Undang-Undang Hukum Dagang.

Pencarian informasi atau yang dikenal dengan sistem temu balik informasi (*Information Retrieval*) bertujuan menghasilkan dokumen yang paling relevan berdasarkan *keyword* pada *query* yang diberikan pengguna. Dokumen dianggap relevan jika suatu dokumen cocok dengan pertanyaan pengguna (Fitri, 2003). Tindakan tersebut mencakup *text indexing*, *inquiry analysis*, dan *relevance analysis*. Ada dua pekerjaan yang ditangani oleh sistem ini, yaitu melakukan *pre-processing* terhadap *database* dan kemudian menerapkan metode tertentu untuk menghitung kedekatan (relevansi atau *similarity*) antara dokumen di dalam *database* yang telah di-*preprocess* dengan *query* pengguna. *Query* yang dimasukkan pengguna dikonversi sesuai aturan tertentu untuk mengekstrak *term-term* penting yang sejalan dengan *term-term* yang sebelumnya telah diekstrak dari dokumen dan menghitung relevansi antara *query* dan dokumen (Karyono, 2012).

Sementara itu, untuk melakukan fungsi pencocokan atau *similarity* digunakan *Vector Space Model*. Pada *algorithm vector space model* digunakan rumus untuk mencari nilai *cosinus* sudut antara dua *vector* dari setiap bobot dokumen (WD) dan bobot dari kata kunci (WK) (Andika, 2015). Metode menghitung kedekatan nilai antar dua dokumen atau kecocokan ini dikenal dengan metode *Cosine Similarity*.

Melihat adanya permasalahan dalam melakukan fungsi pencarian informasi pada Aplikasi Kitab Undang-Undang Hukum Pidana, sementara terdapat beberapa metode yang telah dikembangkan dalam *information retrieval*,

peneliti tertarik untuk melakukan penelitian dan pengembangan dari Aplikasi Kitab Undang-Undang Hukum Pidana ke dalam Aplikasi Kitab Undang-Undang Hukum Dagang dengan judul “Penerapan Metode TF-IDF dan *Cosine Similarity* dalam Aplikasi Kitab Undang-Undang Hukum Dagang”.

1.2 Identifikasi Masalah

Berdasarkan latar belakang permasalahan yang telah dipaparkan, dapat diidentifikasi berbagai permasalahan yang muncul, sehingga perlu untuk dilakukan penelitian, diantaranya :

1. Aplikasi Kitab Undang-Undang Hukum Pidana yang telah ada dalam versi Android sebelumnya, belum mampu menjalankan fungsi dari sistem temu kembali informasi (*Information Retrieval*) dengan baik. Hal ini dapat dilihat dari ketidakmampuan aplikasi dalam memberikan hasil pencarian untuk suatu frase tertentu.
2. Perlu dilakukan pembuktian bahwa salah satu metode yang ada dalam *text mining*, yaitu metode TF-IDF (*Term Frequency-Invers Document Frequency*) yang berfungsi untuk melakukan pembobotan kata, dapat diterapkan dalam suatu sistem temu kembali informasi.
3. Perlu dilakukan pembuktian bahwa salah satu metode yang digunakan untuk fungsi pencocokan, yaitu metode *Cosine Similarity*, dapat diterapkan dalam suatu sistem temu kembali informasi.

4. Perlu dilakukan pembuktian bahwa metode-metode yang digunakan untuk sistem temu kembali informasi dapat diterapkan pada Kitab Undang-Undang Hukum Dagang (KUHD).

1.3 Rumusan Masalah

Melihat permasalahan yang ada, maka perlu diteliti tentang :

1. Bagaimana membuat aplikasi untuk melakukan pencarian indeks pada Kitab Undang-Undang Hukum Dagang ?
2. Bagaimana menguji hasil kinerja sistem dalam melakukan pencarian dengan jumlah *query* 1 hingga 4 kombinasi kata ?
3. Bagaimana menguji kelayakan aplikasi Kitab Undang-Undang Hukum Dagang ?

1.4 Batasan Masalah

Dalam penelitian ini perlu adanya batasan masalah diantaranya :

1. Aplikasi dapat terintegrasi dalam Sistem Operasi *Windows*.
2. Hasil pencarian dari sebuah *keyword* dan frase dapat merujuk pada maksimal 10 pasal yang bersesuaian.
3. *Query* yang diinput maksimal 4 buah kombinasi kata.

1.5 Tujuan Penelitian

Penelitian ini dilakukan untuk memenuhi beberapa tujuan diantaranya :

1. Membuat aplikasi dengan menerapkan metode TF-IDF (*Term Frequency–Invers Document Frequency*) dan *Cosine Similarity* untuk menyusun indeks dari Kitab Undang-Undang Hukum Dagang (KUHD).
2. Mengukur kinerja sistem aplikasi KUHD menggunakan metode Blackbox dan uji *Performance Measure*.
3. Mengukur kelayakan sistem menggunakan uji korelasi Pearson.

1.6 Manfaat Penelitian

Melalui penelitian ini, diharapkan dapat memberikan manfaat :

1. Bagi pihak yang memiliki profesi berkaitan dengan hukum, aplikasi ini dapat menjadi media yang mempermudah dalam melakukan studi kasus.
2. Bagi mahasiswa hukum, aplikasi ini dapat menjadi media pembelajaran dalam melaksanakan kegiatan belajar mandiri.
3. Bagi peneliti, aplikasi ini dapat menjadi referensi untuk mengembangkan aplikasi dan penelitian yang berkaitan dengan *Text Mining*.
4. Bagi masyarakat yang lebih luas, aplikasi ini dapat menjadi sumber informasi yang berkaitan dengan hukum dagang.

1.7 Penegasan Istilah

Agar mendapatkan gambaran yang jelas dan menghindari kekaburan dalam memahami maksud dari penelitian ini, maka akan dijelaskan beberapa istilah yang digunakan yaitu :

- Penerapan menurut Kamus Besar Bahasa Indonesia (KBBI) adalah suatu proses, cara, atau perbuatan menerapkan. Artinya bisa menjadi pemasangan (dalam konteks ilmu teknik) dan pemanfaatan (dalam konteks ilmu sosial).
- Metode menurut Kamus Besar Bahasa Indonesia (KBBI) adalah cara kerja yang bersistem untuk memudahkan pelaksanaan suatu kegiatan guna mencapai tujuan yang ditentukan.
- Aplikasi menurut Indrajani (2009:5) adalah program yang menentukan aktifitas pemrosesan informasi yang dibutuhkan untuk penyelesaian tugas-tugas khusus dari pemakai komputer.

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Kitab Undang-Undang Hukum Dagang

Hukum Dagang adalah hukum yang mengatur tingkah laku manusia yang turut melakukan perdagangan dalam usahanya memperoleh keuntungan. Dapat juga dikatakan, Hukum Dagang ialah hukum yang mengatur hubungan hukum antara manusia-manusia dan badan-badan hukum satu sama lainnya, dalam lapangan perdagangan (Kansil, 2008:20).

Kitab Undang-Undang Hukum Dagang (KUHD) Indonesia kira-kira satu abad yang lalu telah dibawa orang Belanda ke tanah air kita, mula-mula ia hanya berlaku bagi orang-orang Eropa di Indonesia (berdasarkan asas konkordansi). Kemudian juga dinyatakan berlaku bagi orang-orang Timur Asing, akan tetapi tidak berlaku seluruhnya untuk orang-orang Indonesia (hanya bagian-bagian tertentu saja).

KUHD yang mulai berlaku di Indonesia pada 1 Mei 1848 terbagi atas dua kitab dan 23 bab. Kitab I terdiri dari 10 bab dan Kitab II terdiri dari 13 bab (Kansil, 2008:21).

Permasalahan yang muncul adalah mencari bagian pasal maupun ayat yang dilakukan secara manual. Dengan kemajuan teknologi yang sangat pesat pada saat ini dibutuhkan sebuah aplikasi yang dapat mengakomodir kebutuhan setiap pengguna akan sebuah aplikasi pengganti

buku yang mudah dibawa serta dapat digunakan kapanpun dan dimanapun dengan efektif (Ruli, 2013).

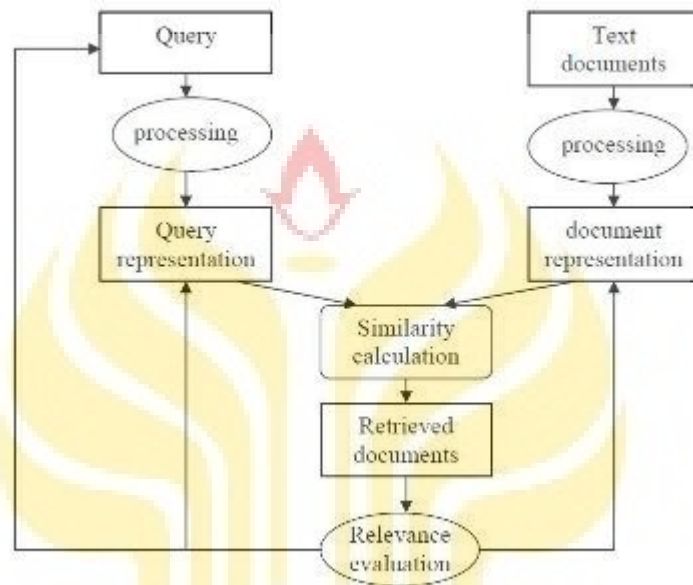
2.1.2 Aplikasi

Aplikasi adalah *software* yang dibeli perusahaan dari tempat pembuatnya. Aplikasi akan menggunakan sistem operasi (SO) komputer dan aplikasi lainnya yang mendukung. Istilah ini mulai perlahan masuk dalam istilah Teknologi Informasi semenjak tahun 1993, yang biasanya juga disingkat dengan app. Secara historis, aplikasi adalah *software* yang dikembangkan oleh sebuah perusahaan (Azmi dalam Ruli, 2013). *Software* aplikasi adalah program yang menentukan aktifitas pemrosesan informasi yang dibutuhkan untuk penyelesaian tugas-tugas khusus dari pemakai komputer (Indrajani, 2009:5).

2.1.3 Konsep Dasar Sistem Temu Balik Informasi

Pencarian informasi atau yang dikenal dengan sistem temu balik informasi (*Information Retrieval*) bertujuan menghasilkan dokumen yang paling relevan berdasarkan *keyword* pada *query* yang diberikan pengguna. Dokumen dianggap relevan jika suatu dokumen cocok dengan pertanyaan pengguna (Fitri, 2003). Tindakan tersebut mencakup *text indexing*, *inquiry analysis*, dan *relevance analysis*. Ada dua pekerjaan yang ditangani oleh sistem ini, yaitu melakukan *pre-processing* terhadap *database* dan kemudian menerapkan metode tertentu untuk menghitung kedekatan (relevansi atau *similarity*) antara dokumen di dalam *database* yang telah di-*preprocess* dengan *query* pengguna. *Query* yang dimasukkan pengguna dikonversi

sesuai aturan tertentu untuk mengekstrak *term-term* penting yang sejalan dengan *term-term* yang sebelumnya telah diekstrak dari dokumen dan menghitung relevansi antara *query* dan dokumen.



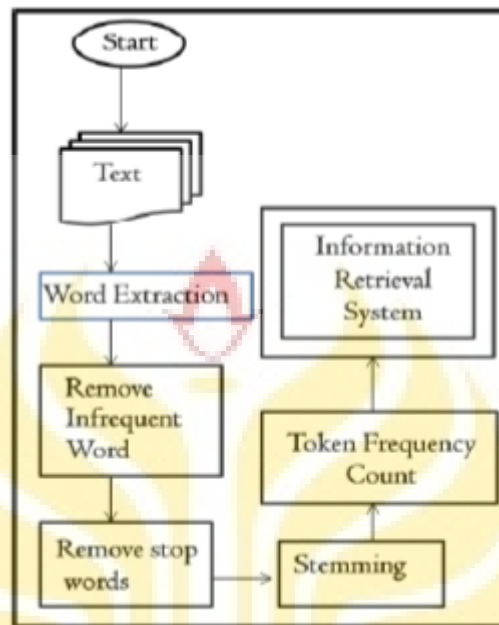
Gambar 2.1 Arsitektur Sistem Temu Balik Informasi

Sebagai hasilnya, sistem mengembalikan suatu daftar dokumen terurut sesuai nilai kemiripannya dengan *query* pengguna. Setiap dokumen (termasuk *query*) direpresentasikan menggunakan model *bag-of-words* yang mengabaikan urutan dari kata-kata di dalam dokumen, struktur sintaktis dari dokumen dan kalimat. Dokumen ditransformasi ke dalam suatu “tas” berisi kata-kata *independen*. *Term* disimpan dalam suatu *database* pencarian khusus yang ditata sebagai sebuah *inverted index*. Index ini merupakan konversi dari dokumen asli yang mengandung sekumpulan kata ke dalam daftar kata yang berasosiasi dengan dokumen terkait dimana kata-kata

tersebut muncul. Proses dalam *Information Retrieval* dapat digambarkan sebagai sebuah proses untuk mendapatkan *retrieve document* dari *collection documents* yang ada melalui pencarian *query* yang diinputkan user (Karyono, 2012).

Sistem temu kembali informasi pada dasarnya dibagi dalam dua komponen utama yaitu sistem pengindeksan (*indexing*) yang menghasilkan basis data sistem dan temu kembali yang merupakan gabungan dari *user interface* dan *look-up-table*. *Indexing* merupakan sebuah proses untuk melakukan pengindeksan terhadap kumpulan dokumen yang akan disediakan sebagai informasi kepada pemakai. Proses pengindeksan bisa secara manual ataupun secara otomatis. Dewasa ini, sistem pengindeksan secara manual mulai digantikan oleh sistem pengindeksan otomatis (Hasibuan, 2015).

Proses *indexing* tidak lepas dari suatu proses yang disebut dengan proses *tokenization*. Proses *Tokenization* merupakan bagian integral dalam sistem IR (*Information Retrieval*), yang masuk dalam *pre-processing* dokumen dan menghasilkan banyak token. Model dari *tokenization* akan ditampilkan pada gambar berikut, tujuan utama dari *tokenization* adalah untuk mengidentifikasi kata/token/konsep dan frekuensi dari masing-masing dokumen yang diinput.



Gambar 2.2 Proses *Tokenization*

Tahap pertama dalam *pre-processing* adalah mengekstraksi seluruh kata di dalam dokumen. Tahap berikutnya, keseluruhan kata yang jarang digunakan akan di hapus, seperti menghapus kata-kata yang jumlahnya kurang dari dua. Hasil dari tahap sebelumnya akan diinput ke dalam tahap berikutnya yaitu tahap penghapusan *stopword*. Tujuan utamanya adalah menghapus kata-kata yang tidak bermanfaat. Definisi umum dari kata-kata tidak bermanfaat ini misalnya kata sifat, konjungtif dan kata-kata lain yang secara umum diabaikan dalam kategori dokumen ini.

Melengkapi *preprocessing*, tahap berikutnya adalah menghitung frekuensi dari masing-masing kata. *Information Retrieval* bekerja pada

hasil dari proses tokenisasi untuk mencapai atau mengolah hasil yang paling relevan untuk ditampilkan (Singh, 2014).

2.1.4 Metode TF-IDF (*Term Frequency – Invers Document Frequency*)

Dalam skema arsitektur temu balik informasi, terdapat sebuah proses yang berkaitan dengan pembobotan kata atau *term* baik secara lokal maupun global. Pembobotan lokal hanya berpedoman pada frekuensi munculnya *term* dalam suatu dokumen dan tidak melihat frekuensi kemunculan *term* tersebut di dalam dokumen lainnya. Pendekatan dalam pembobotan lokal yang paling banyak diterapkan adalah *term frequency* (tf) meskipun terdapat skema lain seperti pembobotan biner, *augmented normalized tf*, logaritmik tf dan logaritmik alternatif.

Pembobotan global digunakan untuk memberikan tekanan terhadap *term* yang mengakibatkan perbedaan dan berdasarkan pada penyebaran dari *term* tertentu di seluruh dokumen. Banyak skema didasarkan pada pertimbangan bahwa semakin jarang suatu *term* muncul di dalam total koleksi maka *term* tersebut menjadi semakin berbeda. Pemanfaatan pembobotan ini dapat menghilangkan kebutuhan *stop word removal* karena *stop word* mempunyai bobot global yang sangat kecil. Namun pada prakteknya lebih baik menghilangkan *stop word* di dalam fase *pre-processing* sehingga semakin sedikit *term* yang harus ditangani. Pendekatan terhadap pembobotan global mencakup *inverse document frequency* (idf), *squared idf*, *probabilistic idf*, GF-idf, entropy. Pendekatan idf merupakan pembobotan yang paling banyak digunakan saat ini. Beberapa aplikasi tidak

melibatkan bobot global, hanya memperhatikan *tf*, yaitu ketika *tf* sangat kecil atau saat diperlukan penekanan terhadap frekuensi term di dalam suatu dokumen (Karyono, 2012). Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen (*database*) (Intan, 2006).

Rumus umum untuk TF-IDF (Al-Talib, 2013) :

$$\text{TF-IDF}(t_i, d_j) = \text{tf}(t_i, d_j) \log N/n_i$$

Keterangan :

- $\text{TF-IDF}(t_i, d_j)$: pembobotan kata atau *term* *i* pada dokumen *j*
- $\text{tf}(t_i, d_j)$: banyak kata atau *term* *i* pada dokumen *j*
- N : total dokumen dalam dataset
- n_i : total dokumen yang memunculkan *term* *i*

2.1.5 Cosine Similarity

Vector space model adalah suatu model yang digunakan untuk mengukur kemiripan antara suatu dokumen dengan suatu *query*. Pada model ini, *query* dan dokumen dianggap sebagai *vektor-vektor* pada ruang *n*-dimensi, dimana *n* adalah jumlah dari seluruh *term* yang ada dalam *leksikon*. *Leksikon* adalah daftar semua *term* yang ada dalam indeks. Salah satu cara untuk mengatasi hal tersebut dalam model *vector space* adalah dengan cara melakukan perluasan vektor. Proses perluasan dapat dilakukan pada *vektor query*, *vektor* dokumen, atau pada kedua *vektor* tersebut.

Pada algoritma *vector space model* digunakan rumus untuk mencari nilai *cosinus* sudut antara dua *vector* dari setiap bobot dokumen (WD) dan bobot dari kata kunci (WK). Rumus yang digunakan dalam *vektor space model* sebagai berikut (Andika,2015) :

$$\text{Cosinus} \rightarrow \text{sim}(d_j, q) = \frac{d_j \cdot q}{|d_j| \cdot |q|} = \frac{\sum_i w_{ij} \cdot w_q}{\sqrt{\sum_i w_{ij}^2} \cdot \sqrt{\sum_i w_q^2}}$$

Keterangan :

- w_{ij} : bobot kata i pada dokumen j
- w_q : bobot *query*

Perhitungan nilai *cosinus* sudut antara dua vektor ini dikenal dengan metode *Cosine Similarity*. Nilai sudut *cosinus* antara dua vektor menentukan kesamaan dua buah objek yang dibandingkan dimana nilai terkecil adalah 0 dan nilai terbesar adalah 1 (Firdaus, 2014).

2.1.6 Evaluasi Sistem Information Retrieval

a. Metode *Blackbox*

Metode *Blackbox* memungkinkan pereayasa perangkat lunak mendapatkan serangkaian kondisi *input* yang sepenuhnya menggunakan semua persyaratan fungsional untuk suatu program.

Blackbox dapat menemukan kesalahan dalam kategori berikut :

- Fungsi-fungsi yang tidak benar atau hilang

- Kesalahan *interface*
- Kesalahan dalam struktur data atau akses basis data eksternal
- Inisialisasi dan kesalahan terminasi
- Validitas fungsional
- Kesensitifan sistem terhadap nilai input tertentu
- Batasan dari suatu data (Rouf, 2012).

b. *Expert Judgement* (Uji Pakar)

Guna mengetahui validitas hasil output dari suatu sistem, maka dibutuhkan sebuah pengujian yang dilakukan oleh orang-orang yang ahli pada bidangnya. Hal ini juga berlaku untuk sebuah sistem *information retrieval* yang dibuat.

Pada *expert judgement* (uji pakar) untuk sistem *information retrieval*, pengujian dilakukan dengan melihat relevansi suatu *query* dengan hasil output sistem. Pengujian relevansi dapat dilakukan dengan menentukan hasil sistem merupakan hasil yang relevan atau tidak relevan. Penentuan juga dapat dilakukan dengan membuat *grade* / tingkatan relevansi hasil, menjadi sangat relevan, cukup relevan dan tidak relevan. Pengujian relevansi sehingga didapat dua kelompok hasil output, yaitu relevan dan tidak relevan, menjadi sangat penting karena dapat digunakan untuk mengevaluasi hasil kinerja sistem (Clough, 2013).

c. *Performance Measure* (Evaluasi Kinerja Sistem)

Metode evaluasi yang biasanya digunakan pada *information retrieval* adalah metode intrinsik dengan *Precision / Recall* dan *F-Measures* yang dirumuskan dengan rumusan (Zaman, 2011) :

$$a. \textit{Recall} = \frac{\textit{jumlah kalimat relevan yang dihasilkan sistem}}{\textit{jumlah kalimat relevan}}$$

$$b. \textit{Precision} = \frac{\textit{jumlah kalimat relevan yang dihasilkan sistem}}{\textit{jumlah kalimat yang dihasilkan sistem}}$$

$$c. \textit{F - Measures} = \frac{2 \times \textit{Recall} \times \textit{Precision}}{\textit{Recall} + \textit{Precision}}$$

d. Uji Korelasi Pearson

Uji korelasi sederhana merupakan suatu Teknik Statistik yang dipergunakan untuk mengukur kekuatan hubungan 2 variable dan juga untuk dapat mengetahui bentuk hubungan antara 2 variabel tersebut dengan hasil yang sifatnya kuantitatif. Kekuatan hubungan antara 2 variabel biasanya disebut dengan Koefisien Korelasi dan dilambangkan dengan symbol “r”. Nilai Koefisien r akan selalu berada di antara -1 sampai +1 ($-1 \leq r \leq +1$).

Pada pengujian korelasi sederhana dapat digunakan suatu rumus sebagai berikut :

(Rumus ini disebut juga dengan Pearson Product Moment)

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}}$$

Keterangan :

n = banyaknya pasangan data x dan y

$\sum x$ = total jumlah dari variabel x

$\sum y$ = total jumlah dari variabel y

$\sum x^2$ = kuadrat dari total jumlah variabel x

$\sum y^2$ = kuadrat dari total jumlah variabel y

Pedoman dalam menentukan tingkat korelasi menggunakan perhitungan ini dapat dilihat dari besarnya nilai r dengan kriteria sebagai berikut :

- $r = 0$, mengisyaratkan tidak adanya korelasi
- $r = 0 - 0.5$, mengisyaratkan korelasi yang lemah
- $r = 0.5 - 0.8$, mengisyaratkan korelasi sedang
- $r = 0.8 - 1$, mengisyaratkan korelasi yang kuat/erat
- $r = 1$, mengisyaratkan korelasi yang sempurna.

2.1.7 Sistem Temu Balik Informasi “Kitab Undang-Undang Hukum Dagang”

Menggunakan Metode TF-IDF dan Cosine Similarity

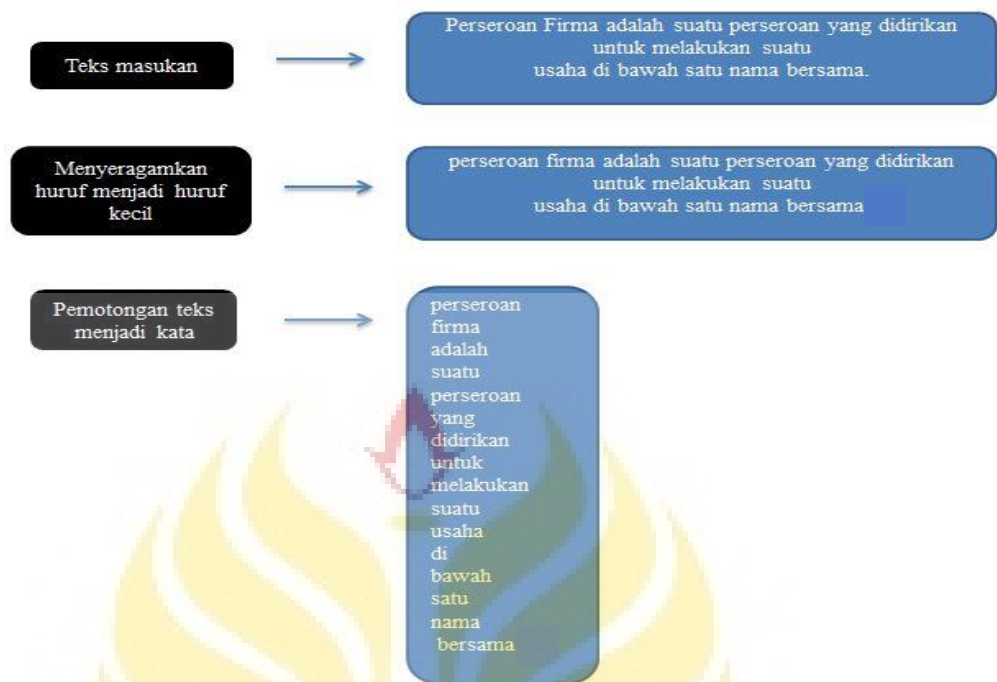
Pada penelitian ini, sistem temu balik Informasi Kitab Undang-Undang Hukum Dagang menggunakan metode TF-IDF untuk melakukan pembobotan kata dari tiap-tiap pasal dan menggunakan *Cosine Similarity* untuk melakukan pencocokan pada suatu frase atau dua buah *query* dengan masing-masing isi pasal yang telah terindeks.

Untuk mendapatkan pasal yang terindeks, masing-masing pasal harus melalui tahap *preprocessing*. Di dalam tahap ini tiap satu pasal dianggap sebagai satu dokumen. Sehingga pada masing-masing pasal akan dilakukan berbagai proses berikut :

1. *Tokenizing*

Tahap *tokenizing* ini memproses tiap dokumen menjadi kata per kata. Selain itu seluruh dokumen akan diseragamkan hurufnya menjadi huruf kecil. Ilustrasi dari tahap ini dijelaskan pada gambar 2.3 berikut.



Gambar 2.3 Tahap *tokenizing* pasal

2. *Filtering*

Tahap *filtering* memfokuskan pada penggunaan *stopword* untuk memperoleh kata-kata penting yang digunakan pada proses berikutnya, seperti yang dijelaskan pada gambar 2.4 berikut.

Gambar 2.4 Tahap *filtering* pasal

Setelah dokumen yang berupa pasal-pasal melalui tahap *preprocessing*, maka setiap dokumen akan dihitung bobotnya. Penghitungan bobot kata ini dilakukan menggunakan metode TF-IDF dengan rumus :

$$TF-IDF(t_i, d_j) = tf(t_i, d_j) \log N/n_i$$

Artinya, masing-masing pasal akan dihitung banyaknya kata yang muncul di dalamnya (*Term Frequency* atau TF), kemudian jumlah keseluruhan dokumen dibagi dengan jumlah dokumen yang memunculkan kata tertentu yang sebelumnya sudah dihitung (N/n_i).

Selanjutnya dihitung kemiripan dari suatu *query* yang diinputkan dengan dokumen menggunakan metode *Cosine Similarity*. Setelah didapatkan hasil yang sesuai dengan *query*, maka akan dimunculkan hasil dokumen (pasal) yang memiliki kemiripan tertinggi.

2.1.8 Penelitian Terdahulu

Beberapa penelitian terdahulu terkait tentang *information retrieval* sudah banyak dirangkum di dalam jurnal-jurnal, baik jurnal nasional maupun jurnal internasional. Begitupun penelitian yang memuat tentang penggunaan metode TF-IDF dan *Cosine Similarity* juga telah banyak dilakukan. Beberapa penelitian yang membahas tentang tema dan metode di atas diantaranya :

Penelitian yang dilakukan oleh Vishal Gupta dan Gurpreet S. Lehal dengan judul “A Survey of Text Mining Techniques and Application”,

disebutkan bahwa dengan menggunakan sejumlah dokumen, fungsi *text mining* akan dijalankan dengan menemukan format dan karakter, serta melakukan *preprocess* yang akan dilanjutkan dengan fase analisis teks hingga informasi dapat diambil maknanya.

Penelitian lain yang dilakukan oleh Rolly Intan dan Andrew Defeng dengan judul “HARD : Subject-based *Search Engine* menggunakan TF-IDF dan Jaccard’s Coefficient”, disebutkan bahwa metode TF-IDF menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan *inverse* frekuensi dokumen yang mengandung kata tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut rendah pada kumpulan dokumen (*database*).

Sementara itu pada penelitian lain yang dilakukan oleh Dhony Syafe’i Harjanto, Sukmawati Nur Endah, dan Nurdin Bahtiar dengan judul “Sistem Temu Kembali Informasi pada Dokumen Teks Menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF)” disebutkan bahwa pembobotan dokumen TF-IDF dapat *retrieve* dokumen sesuai dengan *query* pengguna. Metode pembobotan dokumen TF-IDF tidak selalu memberikan hasil performansi yang baik pada koleksi pengujian, karena tidak semua dokumen *retrieve* merupakan dokumen yang relevan.

Pada penelitian yang dilakukan oleh Juan Ramos yang berjudul “Using TF-IDF to Determine Word Relevance in Document Queries” dijelaskan bahwa TF-IDF merupakan algoritma yang simpel dan efisien untuk digunakan dalam pencocokan kata dalam suatu *query* dengan sejumlah dokumen yang relevan.

Hasil serupa juga dipaparkan dalam penelitian yang dilakukan oleh Diana Permata Sari dan Ayu Purwarianti dengan judul “Ekstraksi Kata Kunci Otomatis untuk Dokumen Bahasa Indonesia Studi Kasus : Artikel Jurnal Ilmiah Koleksi PDII LIPI”. Hasil penelitian tersebut diperoleh bahwa tahapan yang dilakukan untuk mengekstrak kata kunci dokumen berbahasa Indonesia secara otomatis, secara umum adalah tahapan praproses, tahapan translasi, dan tahapan pencocokan kandidat kata kunci dengan daftar kata kunci. Nilai akurasi terbaik yang diperoleh dari hasil keluaran sistem didapat dengan menggunakan teknik pembobotan $TF \times IDF$.

Penelitian yang berkaitan dengan *Cosine Similarity* telah dilakukan oleh Sugiyanto, Bayu Surarso dan Aris Sugiharto dengan judul “Analisa Performa Metode Cosine dan Jacard pada Pengujian Kesamaan Dokumen” dihasilkan bahwa penggunaan pengukuran kemiripan abstrak baik dengan Cosine maupun Jacard secara rata-rata mengindikasikan bahwa keduanya memiliki performa yang tinggi, namun jika dibandingkan terlihat bahwa pengukuran menggunakan *Cosine Similarity* memiliki tingkat akurasi yang lebih baik yaitu sebesar 0,949808 sedangkan Jacard sebesar 0,949077.

Hasil serupa ditunjukkan pada penelitian yang dilakukan oleh Vikas Thada dan Dr. Vivek Jaglan berjudul “Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm”. Penelitian tersebut menyebutkan bahwa nilai hasil kecocokan terbaik dalam eksperimen ini dimiliki oleh *Cosine* kemudian diikuti *Dice* dan *Jaccard*.

2.1.9 Perangkat Pembangun Aplikasi

a. Bahasa Pemrograman PHP

PHP adalah singkatan dari *Hypertext PreProcessors*, *PHP* saat ini adalah Bahasa pemrograman interpreter yang paling banyak digunakan saat ini karena bersifat *open source* dan juga paling banyak didukung oleh banyak *web server*. *PHP* dapat digunakan oleh banyak sistem operasi dari Windows, Linux maupun BSD. *PHP* umumnya diintegrasikan dengan aplikasi database yang juga *opensource* seperti MySQL maupun PostgreSQL, tapi bisa juga diintegrasikan dengan Microsoft SQL, Access maupun Oracle (Manduro, 2011).

b. XAMPP

XAMPP adalah sebuah software *web server* apache yang di dalamnya sudah tersedia database server MySQL dan dapat mendukung pemrograman PHP. XAMPP merupakan software yang mudah digunakan, gratis dan mendukung instalasi di Linux dan Windows. Keuntungan lainnya adalah cuma menginstal satu kali sudah tersedia Apache Web Server, MySQL

Database Server, PHP Support (PHP 4 dan PHP 5) dan beberapa *module* lainnya (Februariyanti, 2012).

c. Adobe Dreamweaver

Adobe Dreamweaver adalah suatu produk web developer yang dikembangkan *Adobe System Inc.*, sebelum produk *Dreamweaver* dikembangkan oleh *Macro-media Inc.* yang kemudian sampai saat ini penembangannya diteruskan oleh *Adobe System Inc.*, *Dreamweaver* dikembangkan dan dirilis dengan kode nama *Creative Suit* (CS) (Sibero dalam Perwira, 2014).

2.2 Kerangka Berfikir

Pada penelitian ini, dirumuskan suatu permasalahan yang menjadi bahan untuk melakukan suatu penelitian, dalam hal ini adalah merancang suatu sistem temu balik informasi yang dapat memproses *query* yang memiliki sejumlah kata tertentu (lebih dari 1 kata), sebagai pembenahan terhadap aplikasi Kitab Undang-Undang Hukum Pidana versi Android yang tidak dapat melakukan fungsi pencarian dengan baik. Agar dapat memecahkan permasalahan di atas, dilakukan studi pustaka, dalam hal ini yang berkaitan dengan *Information Retrieval*. Tujuannya agar didapatkan metode yang dapat diterapkan pada penelitian ini.

Di dalam *information retrieval* dikenal banyak sekali metode yang bertujuan untuk membobot suatu kata dan melakukan fungsi pencocokan. Pada penelitian kali ini diujicobakan metode TF-IDF untuk melakukan fungsi

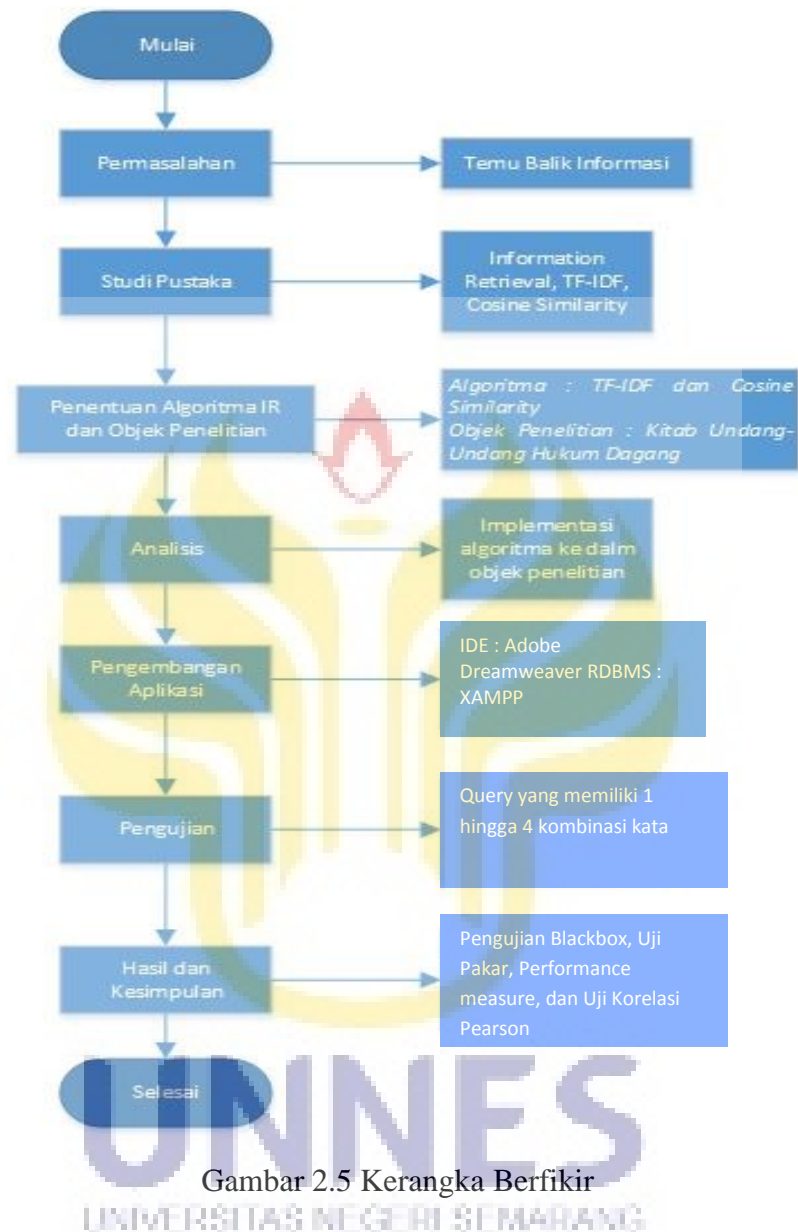
pembobotan kata pada sistem. Sementara untuk melakukan fungsi pencocokan digunakan metode *Cosine Similarity*.

Setelah ditentukan metode yang digunakan untuk penelitian, dibuatlah algoritma TF-IDF dan *Cosine Similarity*. Agar dapat mengetahui proses dan hasil dari kedua metode ini, maka diperlukan suatu objek penelitian yang dapat menerapkan kedua metode tersebut. Objek penelitian yang dipakai merupakan sekumpulan dokumen yang memungkinkan untuk dipakai pada sistem temu balik informasi. Maka dipilihlah Kitab Undang-Undang Hukum Dagang yang merupakan sekumpulan pasal-pasal berupa teks yang berjumlah 984 buah pasal.

Penerapan kedua metode di atas dalam program, disusun menggunakan bahasa pemrograman PHP yang dapat dibuat menggunakan aplikasi *Adobe Dreamweaver* dan menggunakan MySQL pada XAMPP untuk melakukan penyimpanan *document collection* berupa pasal-pasal dari Kitab Undang-Undang Hukum Dagang.

Setelah dilakukan perancangan program, maka dilakukan uji coba menggunakan *query* dengan kombinasi kata dari 1 hingga 4 kata. Untuk dapat melihat kinerja sistem, maka dilakukan analisa menggunakan perhitungan *recall*, *precision* dan *f-measure*. Pengujian *Blackbox* juga dilakukan untuk melihat ada tidaknya *error* pada sistem.

Seluruh proses di atas disusun sebagai sebuah kerangka berfikir dalam penelitian yang akan dilakukan. Penjelasan secara ringkas dapat dilihat pada gambar 2.5 berikut :



BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Aplikasi yang membantu dalam pencarian indeks dalam Kitab Undang-Undang Hukum Dagang menggunakan metode TF-IDF dan Cosine Similarity sudah berhasil dibuat.
2. Tingkat kinerja sistem yang diuji menggunakan perhitungan *Performance Measure* menunjukkan angka keberhasilan *f-measure* 56,93% dalam memproses *query* dengan 1 hingga 4 kombinasi kata.
3. Tingkat korelasi hasil output sistem dengan pasal relevan dalam kitab yang dihitung menggunakan uji korelasi Pearson menghasilkan nilai koefisien korelasi Pearson sebesar 0,71.

5.2 Saran

Berdasarkan hasil penelitian yang dilakukan, berikut beberapa saran untuk penelitian selanjutnya:

1. Perlu adanya penelitian untuk memberikan performa yang lebih baik pada pencarian kata atau frase yang sebenarnya sudah terindeks di dalam Kitab

Undang-Undang Hukum Dagang, agar dapat berada pada urutan teratas saat dilakukan *sorting*.

2. Perlu diterapkan *stemming*, yaitu pemrosesan kata berimbuhan menjadi kata dasar, sehingga proses pembobotan kata dan pencocokan dengan *query* menjadi lebih baik.
3. Perlu dilakukan penelitian lebih lanjut mengenai sistem temu balik informasi dengan metode TFIDF dan *Cosine Similarity* menggunakan data pembelajaran dan pengujian dengan jumlah banyak.
4. Perlu dilakukan penelitian agar *query*, berupa suatu kata umum, yang banyak disebutkan dalam kitab dapat diproses lebih cepat.
5. Perlu dilakukan penelitian dalam hal pembobotan kata untuk kata yang sudah terindeks agar mendapatkan nilai korelasi yang lebih tinggi.

DAFTAR PUSTAKA

- Al-Talib, G. dan Hassan, H. 2013. A Study on Analysis of SMS Classification Using TF-IDF Weighting. *International Journal of Computer Networks and Communications Security* 1(5): 189-194.
- Amin, F. 2012. Sistem Temu Kembali Informasi dengan Metode Vector Space Model. *Jurnal Sistem Informasi Bisnis* (2) : 78-83.
- Andika, Ari. 2015. Perancangan Aplikasi Pengukuran Similaritas pada Dokumen dengan Metode Semantic. *Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI)* V(3): 13-19.
- Brata, D. dan Hetami, A. 2015. Perancangan Information Retrieval (IR) untuk Pencarian Ide Pokok Teks Artikel Berbahasa Inggris dengan Pembobotan Vector Space Model. *Jurnal Ilmiah Teknologi dan Informasi ASIA* 9(1): 53-59.
- Clough, P. dan Sanderson, M. 2013. Evaluating the performance of information retrieval systems using test collections *Information Research*. <http://InformationR.net/ir/18-2/paper582.html>. 27 Januari 2016 (14.31).
- Firdaus, A. dkk. 2014. Aplikasi Pendeteksi Kemiripan pada Dokumen Teks Menggunakan Algoritma Nazief & Adriani dan Metode Cosine Similarity. *Jurnal Teknologi Informasi* 10(1): 96-109.
- Fitri, Meisya. Perancangan Sistem Temu Balik Informasi dengan Metode Pembobotan Kombinasi TF-IDF untuk Pencarian Dokumen Berbahasa Indonesia. <http://jurnal.untan.ac.id/index.php/justin/article/view/1319>. 15 Agustus 2015 (13.42).
- Gupta, V. dan Lehal, G. 2009. A Survey of Text Mining Techniques and Applications. *Journal of Emerging Technologies in Web Intelligence* 1(1): 60-76.
- Harjanto, D. dkk. 2012. Sistem Temu Kembali Informasi pada Dokumen Teks Menggunakan Metode Term Frequency Inverse Document Frequency (TF-IDF). *Jurnal Sains dan Matematika (JSM)* 20(3): 64-70.

- Harnanto, AD. 2013. Aplikasi Pendukung Pembelajaran Bahasa Isyarat pada Anak Berkebutuhan Khusus. http://eprints.ums.ac.id/26127/15/Naskah_Publikasi.pdf. 2 Februari 2016 (12.21)
- Hasibuan, Z. dan Andri, Y. Penerapan Berbagai Teknik Sistem Temu-Kembali Informasi Berbasis Hiperteks. <http://repository.ui.ac.id/dokumen/lihat/6396.pdf>. 21 Oktober 2015 (12.47).
- Indrajani. 2009. *Sistem Basis Data dalam Paket Five In One*. PT. Elex Media Komputindo. Jakarta.
- Intan, R. dan Defeng, A. HARD:Subject-based Search Engine Menggunakan TF-IDF dan Jaccard's Coefficient. <http://jurnalindustri.petra.ac.id/index.php/ind/article/view/16502>. 23 Januari 2016 (22.55).
- Justika Siar Publika. 1999. *Kitab Undang-Undang Hukum Dagang (Wetboek van Koophandle voor Indonesia)*. <http://www.hukumonline.com>. 17 Februari 2015 (18.53).
- Kansil. 2008. *Pokok-Pokok Pengetahuan Hukum Dagang Indonesia*. Sinar Grafika. Jakarta.
- Karyono, G. dan Utomo, F. 2012. Temu Balik Informasi pada Dokumen Teks Berbahasa Indonesia dengan Metode Vector Space Retrieval Model. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2012*. 23 Juni 2012: 283-289.
- Lamusu, M. dan Koniyo, M. 2014. Pembuatan Aplikasi Kamus 3 Bahasa pada Smartphone Android. <http://eprints.ung.ac.id/4555/>. 19 Januari 2015 (12.26).
- Luthfiarta, A. dkk. 2013. Algoritma Latent Semantic Analysis (LSA) pada Peringkat Dokumen Otomatis untuk Proses Clustering Dokumen. *Seminar Nasional Teknologi Informasi dan Komunikasi Terapan 2013 (SEMANTIK 2013)*. 16 November 2013: 13-18.
- Manduro, A. 2011. Pengembangan Sistem Informasi Lapas Narkoba untuk Menunjang Pengungkapan Kasus Narkoba di Lembaga Pemasyarakatan pada Puslitbang dan Info Badan Narkotika Nasional. <http://papers.gunadarma.ac.id/index.php/mmsi/article/download/14868/1413/1>. 27 Oktober 2015 (12.18).

- Menteri Pendidikan Nasional Republik Indonesia. 2008. *Kamus Bahasa Indonesia*. Pusat Bahasa Departemen Pendidikan Nasional. Jakarta.
- Ramos, J. Using TF-IDF to Determine Word Relevance in Document Queries. <http://citeseerx.ist.psu.edu/viewdoc/download>. 7 Mei 2015 (00.08).
- Rouf, A. 2012. Pengujian Perangkat Lunak dengan Menggunakan Metode White Box dan Black Box. *HIMSYATECH* 8(1).
- Rozas, I. dan Sarno, R. 2011. SiPKoKI ISO 27001 : Sistem Pemilihan Kontrol Keamanan Informasi Berbasis ISO 27001. *Seminar Nasional Pascasarjana XI-ITS*. 27 Juli 2011.
- Salton, G. dan Buckley, C. 1988. Term-Wighting Approaches in Automatic Text Retrieval. <http://www.cs.odu.edu/~jbollen/IR04/readings/article1-29-03.pdf>. 10 Desember 2015 (22.30).
- Sari, D. dan Purwarianti, A. 2014. Ekstraksi Kata Kunci Otomatis untuk Dokumen Bahasa Indonesia Studi Kasus : Artikel Jurnal Ilmiah Koleksi PDII LIPI. *BACA: Jurnal Dokumentasi dan Informasi* 35(2): 139-147.
- Sharma, M. dan Patel, R. 2013. A Survey on Information Retrieval Models, Techniques and Applications. *International Journal of Emerging Technology and Advanced Engineering (IJETA)* 3(11): 542-545.
- Singh, V. dan Saini, B. 2014. An Effective Pre-Proceing Algorithm for Information Retrieval Systems. *International Journal of Database Management Systems (IJDMS)* 6(6): 13-24.
- Sugiyamto, dkk. 2014. Analisa Performa Metode Cosine dan Jacard pada Pengujian Kesamaan Dokumen. *Jurnal Masyarakat Informatika* 5(10): 1-8.
- Teknik Elektronika. 2015. Pengertian dan Analisis Korelasi Sederhana dengan Rumus Pearson. <http://www.teknikelektronika.com>. 29 Januari 2016 (11.27).
- Thada, V. dan Jaglan, V. 2013. Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm. *International Journal of Innovations in Engineering and Technology (IJIET)* 2(4): 202-205.

- Zaman, B. dan Winarko, E. 2011. Analisis Fitur Kalimat untuk Peringkas Teks Otomatis pada Bahasa Indonesia. *Indonesia Journal Computing and Cybernetics System (IJCCS)* 5(2): 60-68.

