

APLIKASI PENCARIAN BUKU DENGAN MENGGUNAKAN METODE TF/IDF DAN VECTOR SPACE BERBASIS WEB PADA PERPUSTAKAAN SEKOLAH MENENGAH ATAS NEGERI 2 PANGKALPINANG

Metha Herliani

*Teknik Informatika STMIK ATMA LUHUR PANGKALPINANG
Jl. Jenderal Sudirman Selindung Lama Pangkalpinang Kepulauan Babel
Email: metha_herliani@yahoo.com*

ABSTRACT

As the development of technology, the impact on libraries is that the need for information must also be fast and can allow a user without having to type the exact title of the book you are looking for. Search engine application that the author made the book aims to facilitate users in terms of finding relevant book titles or irrelevant so the user will be able to see the degree of similarity of the titles sought from the results. The method I use in this application is a method of vector space model which in this method the document search results will be sorted based on the weight of the search words contained in the document. One weighting algorithm is an algorithm $tf-idf$ influenced by the frequency of occurrence of the word in a document and the frequency of documents that have the word and text pemprofilan trigram that every character in the query inputted in pieces in the constitution. To determine the relevant level of an inputted document matches the category of the similarity of a document by using Jaccard algorithm. At this stage of measuring the effectiveness of obtaining the information required testing accuracy with precision and testing formulas acquisition by formula recall. Can be inferred using the method the user more easily find the desired book titles without the need to type in a search query exactly the same titles.

Keywords: *library, trigram, vector space model, weighting text, recall, precision, query*

1. Pendahuluan

1.1 Latar Belakang

Perkembangan teknologi saat ini kian hari kian pesat, hampir semua bidang teknologi sekarang berbasis internet. Salah satu teknologi yang saat ini kita rasakan manfaatnya adalah World Wide Web atau dikenal dengan nama Web. Web kini menjadi sumber data yang sangat berharga untuk setiap pengguna karena didalam web kumpulan dokumen – dokumen saling terhubung dan dapat diakses melalui koneksi internet. Setiap individu didunia bebas menambahkan konten dalam web milik mereka masing – masing. Makin banyak informasi yang ditambahkan, makin besar ukuran web dan semakin sulit pula untuk mencari informasi yang benar – benar diinginkan dalam web, oleh karena itu dibutuhkan suatu teknologi untuk mendapatkan informasi yang benar – benar diinginkan tersebut.

Hal ini juga dialami oleh para pengguna informasi dalam memanfaatkan teknologi internet. Dimana web digunakan sebagai sarana penyimpanan dokumen – dokumen online yang kian hari semakin berkembang karena dipandang sebagai peluang potensial bagi dunia akademik untuk menyimpan koleksi yang ada dipergustakaan seperti buku, jurnal ilmiah, majalah dan lain sebagainya. Teknologi yang tepat untuk mengembangkan, membangun dan mengelola koleksi dokumen perpustakaan yang benar – benar

diinginkan pengguna merupakan tujuan kedepan semua pihak.

Tujuan perpustakaan sendiri adalah sebagai pengarsipan yang ada diinstitusi alah untuk memelihara dokumen yang memiliki harga keilmuan yang nantinya dapat diakses dimasa depan sama mudahnya diakses pada saat ini. Dengan semakin kecilnya media data, penyimpanan menggunakan medai digital menjadi sangat menarik. Keuntungan penyimpanan menggunakan medai digital adalah pencarian dapat dilakukan tidak hanya di antara masukan katalog elektronik, tetapi dapat menampilkan sepanjang keseluruhan isi dokumen.

Dewasa ini, bidang pendidikan mengalami kemajuan yang cukup pesat. Institusi pendidikan berskala besar seperti Sekolah Menengah Atas Negeri 2 Pangkalpinang yang bertempat dijalan kalamaya, rt03/rw02, kecamatan Bacang, kelurahan Bukit Intan dengan jumlah jurusan yang lebih dari satu dan jumlah siswa yang cukup banyak, yang telah banyak menghasilkan lulusan dari berbagai jurusan Ilmu Pengetahuan Alam dan Ilmu Pengetahuan Sosial. Tentunya buku – buku yang digunakan sebagai sarana informasi tambahan serta referensi bagi siswa dan guru pengajar disimpan dengan baik dan tersusun rapi pada perpustakaan sekolah.

Dalam hal ini, perpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang

merupakan salah satu wadah yang menyediakan fasilitas dimana siswa dapat mengakses layanan buku dari berbagai jurusan yang ada. Kumpulan buku tersebut akan terus bertambah jumlah dan jenisnya pada tiap semester. Dan pencatatan jumlah buku yang ada hanya mencatat jumlah inventarisnya dan tidak dilakukan juga pencatatan spesifikasi buku seperti buku fiksi ataupun buku non fiksi serta rak – rak penyimpanan buku tidak berdasarkan kode – kode buku standart yang ditetapkan standar penomoran buku pada perpustakaan.

Hal ini mendorong penulis untuk mengadakan penelitian untuk tugas akhir pada perpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang dengan mengangkat judul “**Aplikasi Pencarian Buku Dengan Menggunakan Metode Tf/Idf Dan Vector Space Berbasis Web Pada Perpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang**”.

Dengan adanya penelitian ini, diharapkan dapat mempermudah pengguna untuk mencari buku sesuai dengan judul buku yang ada dimana tingkat keakuratan pencarian judul buku ini yang menjadi pembahasan titik utama pada penelitian ini. Yang nantinya aplikasi ini tentu digunakan untuk mempermudah user mencari buku yang diinginkan hanya dengan mengetikkan kata kunci (keyword) judul buku yang ada dan walaupun list judul buku yang diketikkan nantinya tidak ada maka aplikasi ini akan mencarikan dari yang paling berhubungan dengan keyword yang di-input-kan oleh user sampai dengan yang apling sedikit hubungannya dengan keyword yang di-input-kan, sehingga dapat memudahkan user dalam mencari buku atau melengkap sumber informasi lainnya.

1.2 Rumusan Masalah

Dari latar belakang diatas ada berbagai macam permasalahan yang dapat diidentifikasi, yaitu sebagai berikut :

- Apakah ada cara untuk mengefektifkan waktu dalam hal mencari judul buku – buku yang ada diPerpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang?
- Apakah dengan menggunakan metode TF/IDF dan Vector Space yang digunakan dalam aplikasi ini dapat mengakuratkan waktu pencarian yang ada diperpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang?
- Bagaimana mengimplementasikan pemodelan TF/IDF dan Vector Space dengan menggunakan bahasa pemrograman php untuk pencarian yang lebih spesifik dan efektif?
- Bagaimana mengimplemtasikan aplikasi pencarian menggunakan TF/IDF dan Vector Space dengan bahasa pemrograman php ?
- Apakah dengan menggunakan metode TF/IDF dan Vector Space yang digunakan dalam aplikasi ini nantinya akan membuat nilai

kecocokan judul buku yang dicari sesuai dengan apa yang diinginkan user?

1.3 Batasan Masalah

Untuk batasan masalah yang akan dibahas pada penyusunan penelitian ini adalah sebagai berikut :

- Pada pemodelan TF/IDF dan Vector Space menggunakan pembobotan frekuensi Trigram.
- Pada perhitungan kesetaraan tingkat relevansi data menggunakan fungsi kesetaraan Jaccard.
- Pemodelan TF/IDF dan Vector Space dibangun dengan menggunakan bahasa pemrograman PHP Dreamweaver untuk interface program dan MySQL sebagai database.
- Aplikasi yang dibangun hanya untuk fasilitas pencarian.
- Judul buku yang dijadikan pencarian hanya judul buku – buku yang ada di perpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang saja.
- Aplikasi menggunakan domain lokal (localhost).

1.4 Tujuan Penelitian

Adapun tujuan penulisan dari penyusunan penelitian ini adalah sebagai berikut :

- Untuk dapat mencari judul buku – buku apa saja yang terdapat diperpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang dengan menerapkan metode TF/IDF dan Vector Space pada aplikasi yang digunakan.
- Mengukur keakuratan dari hasil pencarian sistem berdasarkan relevansi dokumen, dari query yang diberikan.
- Sebagai salah satu syarat kelulusan semester 8 yang merupakan tugas akhir perkuliahan / skripsi pada Sekolah Tinggi Ilmu Manajemen dan Informatika Atma Luhur Pangkalpinang.

2. Landasan Teori

2.1 Konsep Dasar Pencarian Informasi

Dalam sistem informasi, terdapat inti dan tujuan, yaitu menghasilkan informasi itu sendiri. Sesederhana apa pun sistem informasi yang dikembangkan, jika bisa menghasilkan informasi yang diharapkan, maka pengembangnya dikatakan berhasil. Namun dilain pihak, secanggih apapun sistem informasi yang dikembangkan, jika tidak dapat menghasilkan informasi yang diharapkan maka pengembangannya dikatakan gagal. Suatu informasi dapat dikatakan berkualitas apabila tiga hal :

- Akurat (Accuracy)

Sebuah informasi dapat dikatakan akurat jika informasi tersebut tidak biasa atau menyesatkan, bebas dari kesalahan –kesalahan dan harus jelas mencerminkan maksudnya.

b. Tepat Waktu (Timeliness)

Informasi yang dihasilkan dari suatu proses pengolahan data yang diberikan harus tepat waktunya, tidak terlambat karena informasi merupakan landasan dalam pengambilan keputusan. Informasi yang terlambat tidak akan mempunyai nilai yang baik, sehingga kalau digunakan sebagai dasar dalam pengambilan keputusan dapat menimbulkan kesalahan dalam tindakan yang diambil. Kebutuhan akan tepat waktunya sebuah informasi itulah yang pada akhirnya akan menyebabkan mahalnya nilai suatu informasi.

c. Relevan (Relevancy)

Informasi akan relevan jika memberikan manfaat bagi pemakainya dan sesuai dengan kebutuhan, serta berkaitan langsung dengan masalah yang ada.

2.2 Pembobotan Istilah

Istilah di dalam suatu indeks harus bisa membedakan kepentingan dari sebuah dokumen pada sebuah informasi. Caranya yaitu dengan pemberian bobot kepada sebuah istilah terhadap suatu dokumen. Semakin tinggi bobot dari sebuah istilah maka semakin penting istilah tersebut dibandingkan dengan istilah lainnya di dalam sebuah dokumen. Bobot dari istilah ini dicantumkan pada inverted index untuk digunakan dalam proses penemu-kembali dokumen.

Pada saat pencarian istilah tunggal digunakan untuk mengidentifikasi dari isi sekumpulan dokumen, pembedaan harus dilakukan antar istilah tunggal berdasarkan perkiraan nilai istilah tersebut sebagai pendeskripsi sebuah dokumen. Hal ini menunjukkan penggunaan dari bobot istilah yang dicantumkan pada saat proses pengidentifikasian (Salton dan Buckley, 1987).

Penggunaan dari bobot istilah selain untuk membedakan kepentingan suatu istilah di dalam sebuah dokumen juga dapat digunakan untuk menggunakan pengurutan saat penemukembali dengan susunan menurun dari bobot yang besar ke kecil sesuai dengan bobot istilah-istilah yang sama antara query dan dokumen.

a. Pemprofilan Text

Pemprofilan Teks (Kategorisasi teks atau klasifikasi teks) adalah sebuah pekerjaan dari bahasa dokumen yang alamiah untuk kategorisasi standar sesuai dengan kontennya [Sebastiani, 2002]. Himpunan Kategori sering disebut "kosa kata terkontrol". Pemprofilan teks atau kategorisasi teks adalah sebuah teknik yang lama dan bisa dikatakan sangat tradisional untuk pencarian informasi di perpustakaan.

Pada penelitian ini, metode yang digunakan untuk melakukan pemprofilan teks ialah dengan menggunakan metode n-gram. Dimana teks akan dibaca kata perkata dan untuk setiap kata akan dibuat n-gram dari kata tersebut. N-gram adalah substrings sepanjang N karakter dari sebuah string

dalam definisi lain n-gram adalah potongan sejumlah n karakter dari sebuah string. N-gram merupakan sebuah metode yang diaplikasikan untuk pembangkitan kata atau karakter. Metode n-gram ini digunakan untuk mengambil potongan-potongan karakter huruf sejumlah n dari sebuah kata yang secara kontinuitas dibaca dari teks sumber hingga akhir dari dokumen. N-gram terdiri dari uni-gram, bi-gram, tri-gram, quad-gram dan quint-gram.

Pada penelitian ini pemotongan n-gram yang digunakan menggunakan tri-gram. Trigram merupakan pembagian potongan query menjadi tiga huruf. Berikut ini contoh tri-gram dalam bentuk representasi dari string "perancangan basis data" adalah {"per", "era", "ran", "anc", "nca", "can", "ang", "nga", "gan", "an_", "n_b", "_ba", "bas", "asi", "sis", "is_", "s_d", "_da", "dat", "ata"}.

Salah-satu keunggulan menggunakan n-gram dan bukan suatu kata utuh secara keseluruhan adalah bahwa n-gram tidak akan terlalu sensitif terhadap kesalahan penulisan yang terdapat pada suatu teks.

b. Metode TF/IDF

Ruang vektor menghitung kemiripan dengan mendefinisikan sebuah vektor yang mempresentasikan setiap dokumen dan sebuah vektor yang merepresentasikan query. Jika kata-kata dalam dokumen dapat direpresentasikan dalam vektor, sangat dimungkinkan untuk membandingkan dokumen dengan query untuk menunjukkan kemiripan konten keduanya. Jika sebuah query dianggap mirip dengan sebuah dokumen, koefisien kemiripan yang mengukur kemiripan antara dokumen dan query dapat dihitung. Dokumen yang kontennya paling mirip dengan konten pada query dianggap paling relevan.

Faktor pembobotan untuk tiap kata dalam dokumen didefinisikan sebagai kombinasi term frequency dan invers document frequency. Untuk menghitung nilai bobot kata, digunakan rumus:

$$W_{in} = \frac{f_{in}}{\log(k_n)}$$

Gambar 1
Formula untuk memberikan frekuensi

Keterangan:

W_{in} = nilai bobot suatu term i terhadap sebuah dokumen n

f_{in} = Nilai frekuensi term i didalam dokumen n

$\log(k_n)$ = Jumlah term dalam dokumen n

2.3 Pengukuran Keakuratan dan Kesetaraan Perolehan Informasi

Pengukuran Keakuratan dan Kesetaraan perolehan informasi ini dilakukan agar bisa didapatkan kesimpulan dari pengujian yang dilakukan terhadap aplikasi dengan menggunakan metode yang ada.

a. Definisi Recall

Pengujian Kelengkapan (Recall) ialah perbandingan jumlah dokumen relevan yang didapatkan sistem dengan jumlah seluruh dokumen relevan yang ada dalam koleksi dokumen (terambil ataupun tak terambil sistem).

Recall:

Jumlah dokumen relevan yang terambil

jumlah dokumen relevan yang ada dalam database

b. Definisi Precision

Pengujian Ketepatan (Precision) ialah perbandingan jumlah dokumen relevan yang didapatkan sistem dengan jumlah seluruh dokumen yang terambil oleh sistem baik relevan maupun tidak relevan.

Precision:

Jumlah dokumen relevan yang terambil

jumlah dokumen yang terambil dalam pencarian

c. Definisi Fungsi Kesetaraan

Kualitas dari sebuah mesin pencari dapat dilihat dari tingkat relevansi data yang didapatkannya. Kebanyakan mesin pencari menggunakan berbagai metode untuk menentukan peringkat hasil pencarian agar mampu memberikan hasil yang maksimal.

Untuk mengetahui similarity atau tingkat kesetaraan antara dokumen yang diinputkan dengan dokumen yang telah ditetapkan kategorinya. Jika similarity dokumen yang diinputkan mendekati similarity dokumen yang telah ditetapkan maka dokumen tersebut adalah dokumen yang cocok untuk masuk kedalam salah satu kategori yang paling mendekati similarity dari dokumen yang telah ditetapkan sebelumnya. Algoritma yang digunakan pada penelitian ini adalah algoritma jaccard dengan menggunakan rumus :

$$sim(Q, D) = \frac{\sum_{j=1}^t w_{qj} * w_{dj}}{\sum_{j=1}^t (w_{qj})^2 + \sum_{j=1}^t (w_{dj})^2 - \sum_{j=1}^t w_{qj} * w_{dj}}$$

Gambar 2

Rumus Algoritma Jaccard

2.4 Model Pengembangan Perangkat Lunak

Metodologi pengembangan perangkat lunak (atau disebut juga model proses atau paradigma rekayasa perangkat lunak) adalah suatu strategi pengembangan yang memadukan proses, metode, dan perangkat (tools) (Modul RPL.pdf). Metode-metode rekayasa perangkat lunak, memberikan teknik untuk membangun perangkat lunak. Berkaitan dengan serangkaian tugas yang luas yang

menyangkut analisis kebutuhan, konstruksi program, desain, pengujian, dan pemeliharaan (Rogers S. Pressman:2002).

3. Pemodelan Proyek

3.1 Objectives Project

Elemen penting dalam mengarahkan sukses tidaknya suatu proyek adalah dalam menentukan objektifitas proyek. Objektifitas proyek dianggap perlu dalam perihal menjadi landasan manajemen proyek dan tujuan akhir didirikannya manajemen proyek itu sendiri.

Membangun Aplikasi Pencarian Buku dengan menggunakan bahasa pemrograman PHP dan database menggunakan MySQL yang akan diimplementasikan pada Sekolah Menengah Atas Negeri 2 Pangkalpinang tersebut. Adapun proyek ini harus selesai dalam waktu tidak lebih dari 43 hari (+ 6 Minggu) dengan membutuhkan biaya tidak lebih dari Rp 30.000.000,-.

3.2 Identifikasi Deliverables

Setelah menentukan objektifitas proyek dan identifikasi atas stakeholder yang terkait atas pengelolaan proyek Aplikasi Pencarian Buku Dengan Menggunakan Metode Tf/Idf Dan Vector Space Berbasis Web Pada Perpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang ini, maka sangatlah penting sebagai bukti riset dan pemodelan proyek disertakan deliverables.

Deliverables merupakan bentuk hasil atau keluaran yang diperkirakan dapat menjadi bukti kerja atas objektifitas proyek. Bentuk dari deliverables dapat dihitung (intangible), atau tidak dapat dihitung (tangible), dapat juga berupa produk barang atau jasa yang dihasilkan setelah pengelolaan proyek.

3.3 Penjadwalan Proyek

Proyek aplikasi pencarian buku dengan menggunakan metode tf/idf dan vector space berbasis web pada perpustakaan sekolah menengah atas negeri 2 pangkalpinang dimulai pada tanggal 3 Maret 2015 dan direncanakan akan selesai pada tanggal 30 April 2015, jam kerja proyek ditentukan sesuai dengan waktu yang telah dijadwalkan yaitu pada hari Senin – Jum'at pukul 08.00 WIB – pukul 12.00 WIB, sedangkan hari Sabtu dan hari Minggu tidak bekerja. Penjadwalan proyek dibuat dengan menggunakan Microsoft Project 2007 dalam format WBS (Work Breakdown Structure), menggunakan Milestone berikut merupakan tahapan dalam penjadwalan proyek Proyek Aplikasi Pencarian Buku Dengan Menggunakan Metode Tf/Idf Dan Vector Space Berbasis Web Pada Perpustakaan Sekolah Menengah Atas Negeri 2 Pangkalpinang.

3.4 Rencana Anggaran Biaya

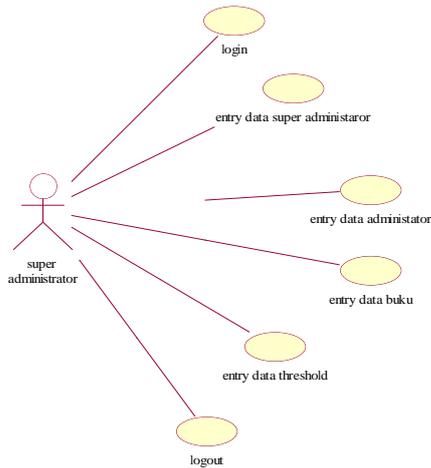
Rancangan Anggaran Biaya adalah perencanaan biaya yang dianggarkan untuk

pengeluaran pengerjaan proyek. Dalam hal ini adalah proyek tugas akhir / skripsi mahasiswa apabila diasumsikan skripsi ini akan dijadikan sebuah proyek. Dalam beberapa model penyusunan RAB dapat langsung disertakan cost of project team dan menghitung satuan standard minimum harga product tertentu.

4. Analisa dan Perancangan Sistem

4.1 Use Case Diagram

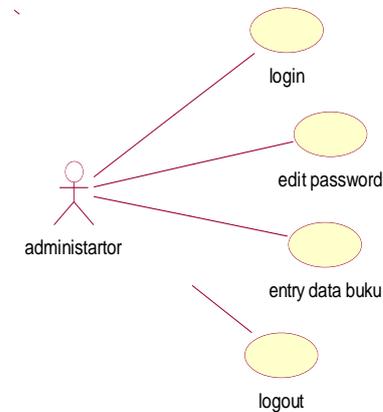
a. Use Case Super Administrator



Gambar 3

Use Case Super Administrator

b. Use Case Diagram Staff



Gambar 4

Use Case Diagram Staff

c. Use Case Diagram Pembaca/User



Gambar 5

Use Case Diagram Pembaca/User

4.2 Rancangan Layar

a. Rancangan Layar Form Login

The screenshot shows a login form for 'PERPUSTAKAAN SEKOLAH MENENGAH ATAS NEGERI 2 PANGKALPINANG SMADA'. It includes a logo box, a header with the school name, and a central area with 'Username' and 'Password' input fields and a 'Login' button.

Gambar 6

Rancangan Layar Form Login

b. Rancangan Layar Form Menu Utama

The screenshot shows the main menu for the library system. It features a header with the school name and a navigation bar with 'Home', 'Administrator', 'Buku', 'Threshold', and 'Logout'. The main content area displays 'Selamat Datang [Super Administrator]' and the school name 'Sekolah Menengah Atas Negeri 2 Pangkalpinang'.

Gambar 7

Rancangan Layar Form Menu Utama

c. Rancangan Layar Form Menu Entry Data Administrator

The screenshot shows the 'Form Entry Data Admin' form. It includes a header with the school name and a navigation bar. The form contains fields for 'Username', 'Nama Lengkap', 'Password', and 'Ulangi Password', along with radio buttons for 'Level' (Super Administrator and Administrator), and 'Simpan' and 'Batal' buttons.

Gambar 8

Rancangan Layar Form Menu entry Data Administrator

d. Rancangan Layar Form Menu Detail Administrator

The screenshot shows the 'Form Detail Admin' form. It features a header with the school name and a navigation bar. The main content area contains a table with columns for 'No', 'Username', 'Nama Lengkap', 'Password', 'Level', and 'Action'. The table is currently empty, with placeholder text '<<tampil>>' in each cell.

No	Username	Nama Lengkap	Password	Level	Action
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>

Gambar 9

Rancangan Layar Form Menu Detail Administrator

e. Rancangan Layar Form Edit Administrator

Gambar 10

Rancangan Layar Form Edit Administrator

f. Rancangan Layar Form Input Buku

Gambar 11

Rancangan Layar Form Input Buku

g. Rancangan Layar Form Edit Buku

Gambar 12

Rancangan Layar Form Edit Buku

h. Rancangan Layar Form Detail Buku

No	Judul Buku	Deskripsi	Kategori	Action
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	edit, hapus, <<tampil>>
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>
<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>	<<tampil>>

Gambar 13

Rancangan Layar Form Detail Buku

i. Rancangan Layar Form Menu Manajemen Threshold

Gambar 14

Rancangan Layar Form Menu Manajemen Threshold

j. Rancangan Layar Form Menu Pencarian

Gambar 15

Rancangan Layar Form Menu Pencarian

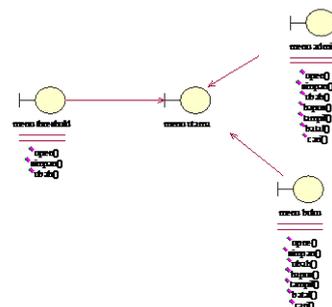
k. Rancangan Layar Form Menu Hasil Pencarian

Gambar 16

Rancangan Layar Form Menu Hasil Pencarian

4.3 Boundary Class

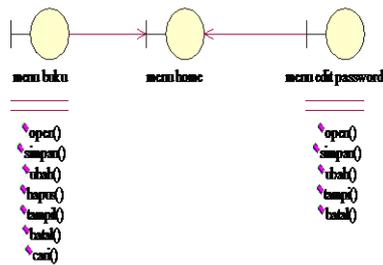
a. Boundary Class



Gambar 17

Gambar Boundary Class Super Administrator

b. Boundary Class Administrator



Gambar 18

Gambar Boundary Class Administrator

4.4 Implementasi dan Uji Coba Aplikasi

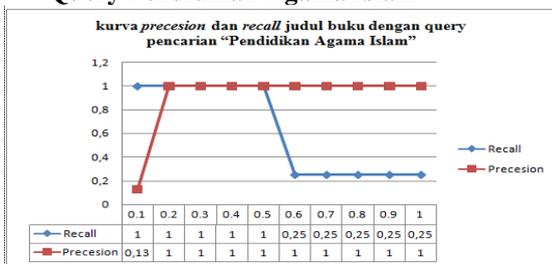
a. Tabel Precision Dan Recall Judul Buku Query Pendidikan Agama Islam

	Nilai Threshold									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Jumlah Dokumen Yang Terambil	30	4	4	4	4	1	1	1	1	1
Jumlah Dokumen Yang Relevan	4	4	4	4	4	1	1	1	1	1
Recall	1	1	1	1	1	0.25	0.25	0.25	0.25	0.25
Precision	0.13	1	1	1	1	1	1	1	1	1

Tabel 1

Tabel Precision dan recall Judul Buku Query Pendidikan agama Islam

b. Kurva Precision Dan Recall Judul Buku Query Pendidikan Agama Islam



Gambar 19

Gambar Kurva Precision Dan Recall Judul Buku Query Pendidikan Agama Islam

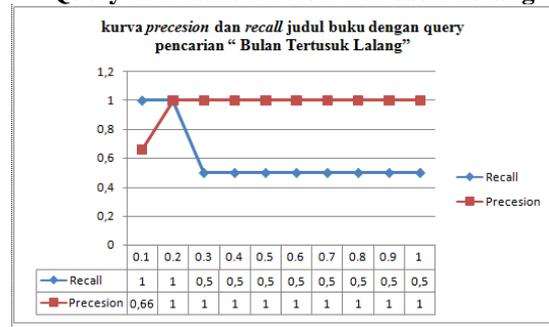
c. Tabel Precision Dan Recall Judul Buku Query Pendidikan Bulan Tertusuk Lalang

	Nilai Threshold									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Jumlah Dokumen Yang Terambil	30	2	1	1	1	1	1	1	1	1
Jumlah Dokumen Yang Relevan	2	2	1	1	1	1	1	1	1	1
Recall	1	1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Precision	0.66	1	1	1	1	1	1	1	1	1

Tabel 2

Kurva Precision Dan Recall Judul Buku Query Pendidikan Agama Islam

d. Kurva Precision Dan Recall Judul Buku Query Pendidikan Bulan Tertusuk Lalang



Gambar 20

Gambar Kurva Precision Dan Recall Judul Buku Query Pendidikan Bulan Tertusuk Lalang

5. Penutup

5.1 Kesimpulan

Kesimpulan yang diambil dari Aplikasi Pencarian Buku Dengan Menggunakan Metode Tf/Idf Dan Vector Space Pada Sekolah Menengah Atas Negeri 2 Pangkalpinang adalah sebagai berikut: Semakin tinggi nilai dari threshold yang ditetapkan maka nilai dari precision (ketepatan) akan semakin tinggi, tetapi nilai dari recall (kelengkapan) akan semakin rendah.

- Penggunaan pemprofilan text menggunakan trigram dapat memberikan kemudahan bagi user yaitu query yang ditulis oleh user tidak mesti sama persis dengan query dokumen yang dicari.
- Keluaran akhir dari aplikasi ini berupa daftar dari buku – buku pelajaran dan umum yang direkomendasikan karena memiliki kemiripan dengan query yang diinputkan oleh user.
- Kecepatan pencarian sebuah informasi tergantung dari jumlah dokumen yang dihasilkan dan jumlah kata kunci yang digunakan sebagai query pencarian.
- Pencarian dengan menggunakan frasa tepat dapat mengurangi nilai recall dari hasil pencarian. Akan tetapi hal ini akan menyebabkan nilai precision menurun, karena tidak semua dokumen yang memiliki informasi yang relevan dapat ditemu – kembalikan.
- Dan juga penggunaan frasa tepat dapat memperlama proses pencarian karena kata kunci harus diproses untuk mencocokkan dokumen.

5.2 Saran

Beberapa hal yang dapat disarankan untuk penelitian dan pengembangan aplikasi pencarian ini nantinya adalah sebagai berikut :

- Perlu dilakukan juga percobaan dengan mencari alternatif pemprofilan text lain misalnya quad-diagram untuk mengetahui kinerja yang baik
- Penggunaan thesaurus (pengelompokan kata yang berhubungan) sehingga dapat

meningkatkan precision dari hasil pencarian judul buku.

- c. Penggunaan model dari aplikasi pencarian yang lainnya untuk dapat membandingkan hasil kinerja aplikasi pencarian sehingga dapat ditemukan model yang paling baik dari aplikasi pencarian.

DAFTAR PUSTAKA

1. Salton, G. & Buckley, C. Term Weighting Approaches in Automatic Text Retrieval. Technical Report No. 87-881. Departement of Computer Science Cornell University Ithaca. New York. 1987.
2. Pressman, Roger S. "Rekayasa Perangkat Lunak (Pendekatan Praktis)." Yogyakarta : Andi. 2002.