

APLIKASI KATALOG PERPUSTAKAAN SMK TANGERANG GLOBAL DENGAN METODE *STEMMER* TALA DAN FUNGSI *JACCARD*

Annisa Rahmala¹⁾, Pipin Farida Ariyani, S.Kom., M.T.I.²⁾

¹Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

^{1,2}Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260

E-mail : 1411501545@student.budiluhur.com¹⁾, pipin.ariyani@gmail.com²⁾

ABSTRAK

Peran petugas perpustakaan yang mengelola seluruh koleksi perpustakaan mengalami kesulitan dalam mengatur koleksi buku-buku tersebut. Semakin tersedia banyaknya koleksi buku dengan kategori yang bermacam-macam dan informasi mengenai buku itu sendiri. Meningkat pula para pustakawan yang membutuhkan informasi mengenai koleksi buku yang ada di perpustakaan. Karena kondisi ini dibutuhkan sebuah sistem yang dapat memberikan informasi. Saat ini perpustakaan SMK Tangerang Global telah memiliki koleksi sekitar 300 buku yang tersedia sebagai penunjang aktivitas belajar para siswa yang dalam kurun waktu 1 tahun kedepan akan ditambahkan sekitar 500 buku. Semakin banyak dan beragamnya buku yang tersedia, maka informasi yang berkualitas secara cepat pun dibutuhkan dalam pencarian koleksi buku di perpustakaan. Adapun informasi yang berkualitas dipengaruhi oleh relevansi, keakuratan dan ketepatan waktu. Seringkali pada web, dimana dalam proses pencarian informasi tertentu, banyak hal yang penting justru terlewatkan dan yang tidak penting banyak terserap. Oleh karena itu dibuatlah aplikasi katalog perpustakaan untuk mencari informasi mengenai koleksi buku yang ada di perpustakaan dengan menggunakan bahasa pemrograman PHP dan basis data MySQL. Dalam penelitian ini dokumen bahasa Indonesia adalah bahasa yang digunakan sebagai sumber acuan. Untuk kebutuhan pencarian informasi yang relevan dari suatu kumpulan informasi diperlukan proses *stemming*, pembobotan, dan pengelompokan dalam pencarian dokumen. Dalam penelitian ini menggunakan salah satu metode yaitu metode *stemming* Tala. Pada sistem pencarian dokumen ini juga menggunakan metode *Logaritmik Term Frequency* berbasis *Vector Space Model* untuk menentukan bobot kata pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut. Semakin besar jumlah kemunculan suatu kata dalam dokumen, semakin besar pula bobotnya dalam dokumen atau akan memberikan nilai kesesuaian yang semakin besar. Disamping itu terdapat fungsi *Jaccard* untuk menghitung tingkat kesamaan (*similarity*) antar *query* dan *term* pada dokumen yang berpengaruh pada pemeringkatan hasil kategorisasi dokumen yang relevan.

Kata kunci : Pencarian Informasi, *Stemming* Tala, *Similarity*, Bobot Kata, *Logaritmik Term Frequency*

1. PENDAHULUAN

Semakin banyak dan beragamnya buku yang tersedia, tanpa adanya sistem penelusuran katalog yang terkomputerisasi maka pengunjung perpustakaan mengalami kesulitan untuk mengetahui ketersediaan bahan pustaka. Berdasarkan permasalahan tersebut timbul sebuah gagasan penulis melakukan riset pada perpustakaan SMK Tangerang Global untuk membangun sebuah mesin pencari yang akan diterapkan pada aplikasi katalog perpustakaan. Pada SMK Tangerang Global, sistem perpustakaan masih dilakukan secara manual dengan melihat daftar judul buku pada data buku perpustakaan tanpa bantuan komputer. Sistem manual tersebut menyebabkan kegiatan operasional perpustakaan berjalan lambat. Dalam merealisasinya penulis akan membuat aplikasi katalog perpustakaan menggunakan bahasa pemrograman php beserta *database* MySQL yang digunakan dalam penyimpanan data.

Aplikasi katalog perpustakaan ini merupakan alat bantu yang berfungsi untuk menunjukkan

ketersediaan koleksi buku yang dimiliki oleh suatu perpustakaan. Selain itu tujuan penelitian yaitu mengimplementasikan metode *Tala* kedalam sistem penelusuran katalog perpustakaan berbasis *web* dan menentukan proses penelusuran katalog perpustakaan berdasarkan *input term* atau kata kunci dari *user*. Dalam penelusuran informasi, jumlah dokumen relevan yang ditelusuri akan dipengaruhi oleh jumlah kata kunci yang digunakan. Kemudian juga terdapat pembobotan sebuah kata menggunakan metode *Logaritmik Term Frequency*. Tahap terakhir yaitu pencarian *similarity* antar dokumen dan *query* menggunakan metode *Jaccard*. Adapun pembobotan yang dilakukan berbasis *Vector Space Model* merupakan model *information retrieval system* yang merepresentasikan dokumen dan *query* dalam bentuk vektor dimensional [1]. Adapun batasan dari aplikasi ini yaitu hanya dapat melakukan pencarian dokumen dengan teks bahasa Indonesia. Selain itu dokumen yang tersedia pada aplikasi katalog harus berformat *pdf* karena jika tidak berformat *pdf* tidak dapat

melakukan pembobotan kata dan aplikasi ini berjalan pada *server localhost*.

2. METODE PENELITIAN

Dalam penyusunan ini penulis melakukan penelitian untuk mengumpulkan data atau informasi yang diperlukan. Adapun metode yang digunakan sebagai berikut :

- a. Penelitian lapangan, yaitu penelitian langsung pada lokasi penelitian.
- b. Metode wawancara, yaitu proses tanya jawab langsung dan sistematis kepada orang yang mengetahui tentang permasalahan yang sedang diamati untuk meyakinkan hal-hal kegiatan observasi yang telah dilakukan.
- c. Penelitian kepustakaan, yaitu penelitian yang dilakukan dengan cara mempelajari buku-buku, literatur-literatur, jurnal, sumber bacaan lainnya yang berkaitan erat hubungannya dengan pembahasan laporan ini.
- d. Metode pengembangan aplikasi yang digunakan dalam mengembangkan aplikasi ini adalah *Waterfall*. Tahap-tahap yang harus di selesaikan secara berurut, yaitu :
 - 1) *Analysis* : proses pencarian data yang berkaitan dengan pembangunan aplikasi katalog perpustakaan seperti jumlah buku yang terdapat di perpustakaan, kategori buku yang tersedia, dan pengguna sistem.
 - 2) *Design* : proses ini digunakan untuk membuat *mockup* rancangan layar, rancangan program, alur program (*flowchart*), dan rancangan *database*.
 - 3) *Coding* : proses membangun program menggunakan bahasa pemrograman PHP dengan *text editor* notepad dan pembuatan *database MySQL*.
 - 4) *Testing / Verification* : proses uji coba program yang telah dibuat akan di *run* di *web server localhost*.

2.1. Proses Information Retrieval

Information retrieval berfokus pada proses yang terlibat di dalam representasi, media penyimpanan, mencari dan menemukan informasi yang relevan dari informasi yang diinginkan oleh *user* [2]. Berikut proses retrieval :

a. Clustering

Clustering dokumen adalah proses mengelompokkan dokumen berdasarkan kemiripan antara satu dokumen dengan dokumen yang lain dalam satu *cluster*. Tujuan *clustering* dokumen adalah untuk memisahkan dokumen yang relevan dari dokumen yang tidak relevan [3].

b. Stemming

Proses *stemming* adalah salah satu cara yang digunakan untuk meningkatkan performa *Information Retrieval* dengan cara mentransformasi kata-kata dalam sebuah teks dokumen ke bentuk kata

dasarnya [4]. Proses dengan menghilangkan semua imbuhan baik yang terdiri dari awalan (*prefixes*), sisipan (*infixes*), akhiran (*suffixes*) dan *confixes* (kombinasi dari awalan dan akhiran) pada kata turunan [3].

c. Tokenizing

Tokenizing adalah pemotongan *string input* berdasarkan tiap kata yang menyusunnya. Pemecahan kalimat menjadi kata-kata tunggal dilakukan dengan men-*scan* kalimat dengan pemisah *whitespace* [4].

d. Filtering

Filtering disebut juga proses pembuangan *stopword* yang dimaksudkan untuk mengetahui suatu kata masuk kedalam *stopword* atau tidak. Pembuangan *stopword* adalah proses pembuangan *term* yang tidak memiliki arti atau tidak relevan [4].

2.2. Metode Pengembangan Sistem

a. Stemmer Tala

Stemmer Tala merupakan adopsi dari algoritma *stemmer* bahasa Inggris terkenal yaitu *Stemmer Porter*. *Stemmer* ini menggunakan *rule base* analisis untuk mencari *root* (kata dasar) sebuah kata. Adapun struktur pembentukan kata dalam *stemmer* bahasa Indonesia Tala adalah sebagai berikut :

- 1) Proses Menghilangkan Partikel
- 2) Proses Menghilangkan Kata Sandang & Kepunyaan
- 3) Menghilangkan Awalan 1
- 4) Menghilangkan Awalan 2
- 5) Menghilangkan Akhiran

b. Vector Space Model

Vector Space Model (VSM) adalah salah satu model matematika yang digunakan pada sistem temu kembali informasi untuk menentukan bahwa sebuah dokumen itu relevan terhadap sebuah informasi. Model ini akan menghitung derajat kesamaan antara setiap dokumen yang disimpan di dalam sistem dengan *query* yang dimasukkan.

c. Pembobotan Term

Term Frequency yaitu faktor yang menentukan bobot *term* pada suatu dokumen berdasarkan jumlah kemunculannya dalam dokumen tersebut [1].

d. Fungsi Similaritas

Jaccard merupakan metode yang digunakan untuk menghitung tingkat kesamaan (*similarity*) antar dua buah objek. Adapun persamaannya sebagai berikut [3] :

$$\text{Sim}(x, y) = \frac{\sum_{i=1}^i x_i y_i}{\sum_{i=1}^i x_i^2 + \sum_{i=1}^i y_i^2 - \sum_{i=1}^i x_i y_i}$$

Keterangan :

Sim(X, Y) = ukuran kesamaan antara *query* dan dokumen

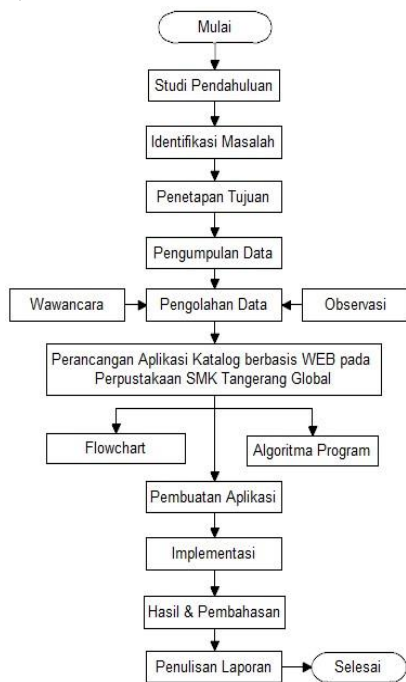
X_i = bobot *term query* inputan
 Y_i = bobot *term query* dokumen

3. RANCANGAN SISTEM DAN APLIKASI

Aplikasi katalog perpustakaan ini dirancang menggunakan pemrograman php yang berjalan di *web browser localhost* dengan berbagai fitur yang tersedia yaitu *upload file*, *download file*, kolom pencarian koleksi pustaka, dan menu yang tersedia yaitu *add/change staff*, *change password admin*, melihat daftar *ebook*. Selain itu terdapat *menu help* untuk memberikan informasi mengenai penggunaan katalog perpustakaan.

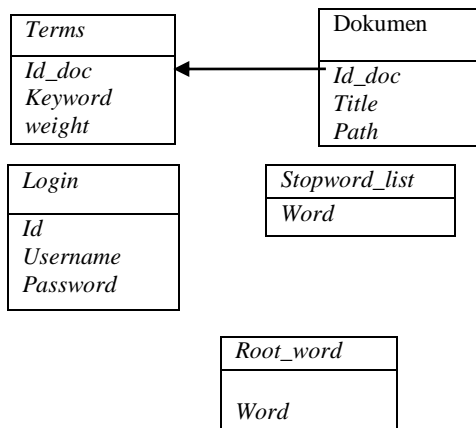
3.1. Desain Alur Penelitian

Berikut adalah alur proses penelitian yang dilakukan penulis dalam penyusunan *paper* dan program.



Gambar 1 : Alur penelitian

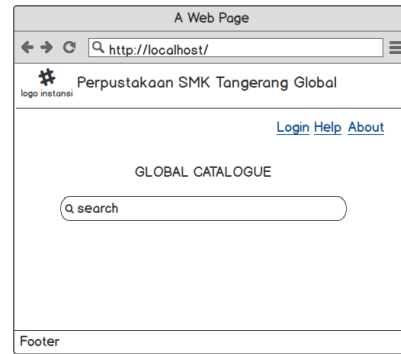
3.2. Rancangan Basis Data



Gambar 2 : LRS (Logical Relational Structure)

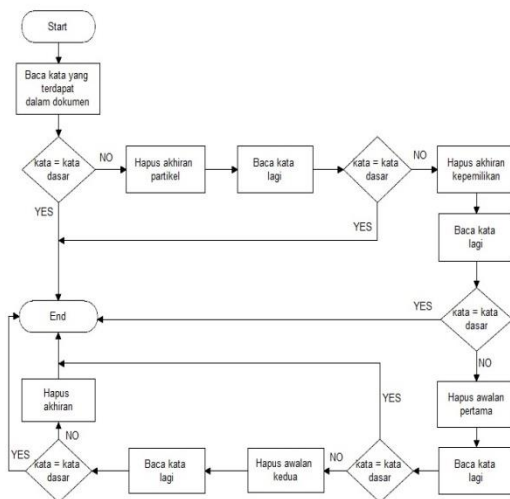
3.3. Rancangan Layar Halaman Utama

Halaman ini adalah halaman utama pencarian dokumen dimana pengguna dapat mencari buku atau artikel yang ingin dicari. Kata kunci bisa dimasukkan ke dalam kolom pencarian *search* kemudian tekan *Enter*, terlihat tampilan halaman *search* pada gambar 2 dibawah ini :



Gambar 3 : Halaman utama pencarian katalog

3.4. Flowchart Proses Stemming



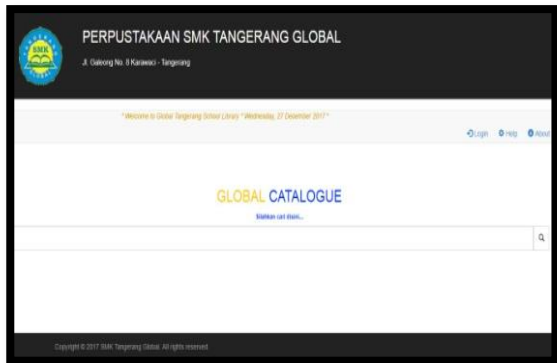
4. HASIL DAN PEMBAHASAN

4.1. Tampilan Layar Aplikasi

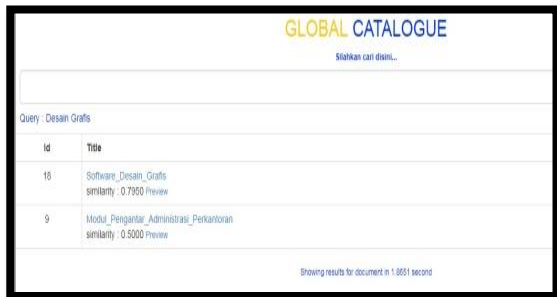
Berikut adalah beberapa tampilan layar yang terdapat pada aplikasi katalog perpustakaan diantaranya halaman utama untuk pencarian dokumen. Halaman ini digunakan secara umum oleh siswa, guru, dan karyawan sekolah. Kemudian ada *form login* yang digunakan untuk masuk ke halaman *dashboard*. Akses halaman *dashboard* hanya diperuntukkan untuk *user* dengan status *admin* dan *staff*. Adapun menu yang tersedia pada halaman *dashboard admin* yaitu *change password admin*, *add/change password staff*, *upload file*, *list book/article*, dan kolom pencarian yang dapat diakses dari halaman *dashboard*. Sementara menu yang tersedia pada halaman *dashboard staff* yaitu *upload file*, *list book/article*, dan kolom pencarian. Sedangkan siswa sekolah hanya diberikan hak akses untuk *download file*.

a. Tampilan Layar Home

Halaman utama (*home*) yaitu fitur utama dari aplikasi katalog perpustakaan dimana pengguna dapat mencari koleksi buku yang tersedia pada perpustakaan.



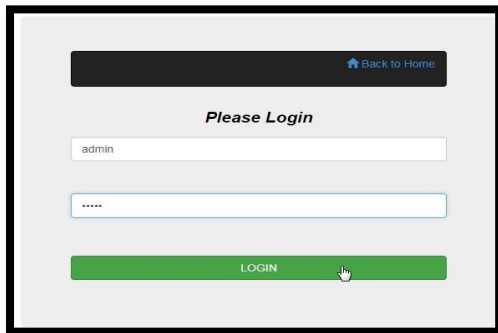
Gambar 4 : Halaman Utama (Home)



Gambar 5 : Tampilan saat searching file

b. Tampilan Layar Form Login

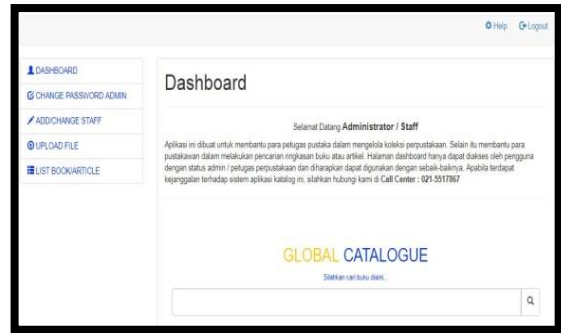
Halaman *login* ini digunakan untuk masuk kedalam halaman *dashboard*. Halaman *login* hanya diperuntukkan bagi pengguna yang sudah mempunyai akun yang telah terdaftar pada *database*.



Gambar 6 : Halaman Login

c. Tampilan Layar Dashboard

Halaman ini hanya dapat diakses oleh pengguna dengan status *admin*. Di halaman *dashboard* ini memiliki beberapa menu diantaranya *change password admin*, *add/change staff*, *upload file*, *list book/article*. Disamping terdapat fitur *searching* di halaman awal *dashboard* yang digunakan untuk mencari buku dari halaman *dashboard admin*.



Gambar 7 : Dashboard admin

Adapun uji coba performa aplikasi katalog perpustakaan ini menggunakan beberapa *sample data* diantaranya terdiri dari 100 *file pdf* yang terbagi menjadi 7 kategori *file* yaitu ringkasan 14 *file* kebudayaan, 16 *file* administrasi perkantoran, 13 *file* TKJ, 10 *file* desain, 13 *file* sains, 18 *file* kesehatan, dan 16 *file* teknologi. Adapun *query* yang digunakan antara lain :

- 1) Modul Administrasi Perkantoran
- 2) Materi Desain
- 3) Tips Menjaga Kesehatan

4.2. Pengukuran Kinerja Aplikasi

a. Hasil Pencarian Dengan Query “Modul Administrasi Perkantoran.”

Tabel 1 : Hasil pencarian *query* “modul administrasi perkantoran”

No	Nama Dokumen	Similarity	Relevan
1	Modul_Tata_Kearsipan	0.8434	Ya
2	Manajemen Perkantoran Modern	0.7884	Ya
3	Pengertian_Router_Dalam Jaringan_Komputer	0.6667	Tidak
4	Proses Penerapan Manajemen Administrasi Perkantoran	0.6445	Ya
5	Otomatisasi Tata Kelola Keuangan	0.6309	Ya
6	Silabus Sistem Operasi X	0.5998	Tidak
7	Otomatisasi Perkantoran	0.5815	Ya
8	Administrasi Pendidikan Dan Teknologi Informasi	0.5115	Ya
9	Pengantar Administrasi Perkantoran II	0.5113	Ya
10	Modul Pengantar Administrasi Perkantoran	0.5021	Ya
11	Otomatisasi Tata Kelola Humas Dan Keprotokolan	0.4057	Ya
12	Teknologi Perkantoran	0.4047	Ya
13	Tata Kelola Perkantoran	0.4	Ya
14	Silabus Simulasi Digital	0.3948	Tidak
15	Administrasi Kepegawaian	0.3931	Ya
16	Silabus Sistem Operasi Jaringan XI	0.3853	Tidak
17	Sejarah Facebook	0.3836	Tidak

18	Perkembangan Teknologi Prosesor	0.3836	Tidak
19	Perusahaan Microsoft	0.3333	Tidak
20	Tutorial Pembuatan Robot Cerdas	0.3333	Tidak
21	Prinsip Tata Ruang Kantor	0.3333	Ya

Tabel 2 : Precision dan recall data dari query “modul administrasi perkantoran”

Jumlah dokumen yang terambil	21
Jumlah dokumen relevan	13
Precision	13/21 = 0.62
Recall	13/16 = 0.81
Waktu pencarian	6.6313 detik

b. Hasil Pencarian Dengan Query “Materi Desain”

Tabel 3 : Hasil pencarian query “materi desain”

No	Nama Dokumen	Similarity	Relevan
1	Desain Komunikasi Visual	0.9349	Ya
2	Modul Pengantar Administrasi Perkantoran	0.7993	Tidak
3	Desain Multimedia Interaktif	0.7620	Ya
4	Silabus Pemrograman Web	0.7131	Ya
5	8 Penemuan Sains Yang Membuka Mata Dunia	0.5461	Tidak
6	Teknik Animasi	0.5461	Ya
7	Modul Tata Kearsipan	0.5440	Tidak
8	Pemrograman Web	0.5440	Ya
9	Sistem Operasi	0.5440	Tidak
10	Software Desain Grafis	0.5404	Ya
11	Otomatisasi Tata Kelola Humas Dan Keprotokolan	0.5330	Tidak
12	Otomatisasi Tata Kelola Keuangan	0.5255	Tidak
13	Pengantar Administrasi Perkantoran II	0.5118	Tidak
14	Mengenal Desain Grafis	0.5000	Ya
15	Pengembangan Produk Kreatif	0.5000	Ya
16	Perkembangan Teknologi Prosesor	0.5000	Tidak
17	Administrasi Pendidikan Dan Teknologi Informasi	0.5000	Tidak

18	Silabus Jaringan Nirkabel	0.4899	Tidak
19	Silabus Sistem Operasi X	0.4854	Tidak
20	Silabus Sistem Operasi XI	0.4854	Tidak
21	Silabus Jaringan Dasar	0.4773	Tidak
22	Silabus Komputer Terapan	0.4668	Tidak
23	Silabus Komunikasi Data	0.4668	Tidak
24	Silabus Simulasi Digital	0.4637	Ya
25	Silabus Perakitan Komputer	0.4553	Tidak

Tabel 4 : Precision dan recall data dari query “materi desain”

Jumlah dokumen yang terambil	25
Jumlah dokumen relevan	9
Precision	9/25 = 0.36
Recall	9/10 = 0.9
Waktu pencarian	4.1492 ik

c. Hasil Pencarian Dengan Query “Tips Menjaga Kesehatan”

Tabel 5 : Hasil pencarian query “tips menjaga kesehatan”

No	Nama Dokumen	Similarity	Relevan
1	Tips Cara Hidup Sehat	0.7683	Ya
2	Modul Pengantar Administrasi Perkantoran	0.6667	Tidak
3	Manfaat Kopi Untuk Kulit	0.6667	Ya
4	Lima Manfaat Ikan Teri	0.6311	Ya
5	Wabah Difteri Di Indonesia	0.6073	Ya
6	Imunisasi Sebagai Wujud Paradigma Sehat	0.4041	Ya
7	Otomatisasi Tata Kelola Humas Dan Keprotokolan	0.4041	Tidak
8	Bahaya Meninggalkan Sarapan Pagi	0.4041	Ya
9	Kandungan Ajaib Air Zamzam	0.4041	Tidak
10	Efek Samping Mengonsumsi Teh Hijau Berlebihan	0.3987	Ya
11	Aturan Imunisasi Difteri Untuk Orang Dewasa	0.3836	Ya

12	Bahaya Makanan Pedas Bagi Kesehatan	0.3836	Ya
13	Pengantar Administrasi Perkantoran II	0.3836	Tidak
14	Antioksidan Senyawa Ajaib	0.3333	Ya
15	Demam Berdarah Masih Mewabah	0.3333	Ya
16	Makanan Protein Tinggi	0.3333	Ya
17	Mengenal Diabetes Mellitus	0.3333	Ya
18	Kaitan Antara Makanan Dengan Mood	0.3333	Ya
19	Ritual Mistis Suku Kengge Di Flores	0.3333	Tidak
20	Robot Yang Mampu Mengenali Wajah	0.3333	Tidak
21	Administrasi Kepegawaian	0.3333	Tidak
22	Pengantar Ekonomi Dan Bisnis	0.3333	Tidak

Tabel 6 : Precision dan recall data dari query “tips menjaga kesehatan”

Jumlah dokumen yang terambil	22
Jumlah dokumen relevan	14
Precision	$14/22 = 0.64$
Recall	$14/18 = 0.78$
Waktu pencarian	6.2084 detik

5. KESIMPULAN

Berdasarkan analisis yang telah dilakukan terhadap permasalahan pada SMK Tangerang Global dari aplikasi yang telah dibuat, maka dapat ditarik beberapa kesimpulan dan saran yang mungkin diperlukan untuk pengembangan sistem ke tahap yang lebih kompleks.

5.1. Kesimpulan

Sesuai dengan pembahasan mengenai penerapan mesin pencari pada katalog perpustakaan menggunakan algoritma *stemmer* Tala dan pemprofilan teks *berbasis vector space model* dengan fungsi *Jaccard* serta pembobotan dengan fungsi *logaritmik term frequency*, maka dapat disimpulkan sebagai berikut :

- Pencarian dokumen yang dilakukan bisa berdasarkan isi *content* dokumen sehingga *user* tidak hanya mengandalkan judul *ebook/artikel* saja sebagai *keyword* dalam proses pencarian.
- Dari hasil uji coba proses perhitungan waktu, didapat bahwa rata-rata waktu tanggap program adalah 5.66 detik.
- Hasil *recall* terhadap pencarian dokumen yang didapat yaitu rata-rata 83%.

5.2. Saran

- Similaritas saat proses pencarian dibuat lebih baik sehingga dalam melakukan searching tidak ada lagi dokumen yang tidak relevan dalam hasil pencarian.
- Dalam melakukan proses pencarian, terdapat pengecekan dan perbaikan pada saat *user* melakukan kesalahan dalam pengetikan *query/kata kunci*.
- Proses *precision* harus lebih tinggi dari 54% untuk meningkatkan kinerja program yang lebih baik.

6. DAFTAR PUSTAKA

- Utomo, M. S. 2013. *Implementasi Stemmer Tala pada Aplikasi Berbasis Web*. Jurnal Teknologi Informasi DINAMIK, 18(1), hal. 41-45.
- Dwi, M. dan Mustafa, Lis Diana. 2017. *Information Retrieval Tugas Akhir dan Perhitungan Kemiripan*. Jurnal Simetris, 8(1), hal. 355-362.
- Sugiyanto, B. S. dan A. S. 2014. *Analisa Performa Metode Cosine dan Jaccard pada Pengujian Kesamaan Dokumen*. Jurnal Masyarakat Informatika, 5(10), hal. 1-8.
- Susandi, D. dan Sholahudin, U. 2016. *Pemanfaatan Vector Space Model pada Penerapan Algoritma Nazief Adriani, KNN dan Fungsi Similarity Cosine untuk Pembobotan IDF dan WIDF pada Prototipe Sistem Klasifikasi Teks Bahasa Indonesia*. Jurnal ProTekInfo, 3(1), hal. 22-29.