ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012

PENGEMBANGAN APLIKASI *WEB BASED DOCUMENTS SIMILARITY MEASURE* MENGGUNAKAN MODEL RUANG VEKTOR PADA DOKUMEN BERBAHASA INDONESIA

Oleh

Made Satria Wibawa, 0815051003
Jurusan Pendidikan Teknik Informatika
Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha
Email : mdsatria@live.com

ABSTRAK

Penelitian ini bertujuan untuk merancang dan membangun sebuah aplikasi berbasis web yang mengimplementasikan model ruang vektor. Metode yang digunakan untuk *pre-processing* dokumen dalam penelitian ini adalah *tokenisation, stopword removal* dan *enhanced confix stemmer* sedangkan untuk perhitungan kesamaan, metode yang digunakan adalah model ruang vektor. Implementasi penelitian ini menghasilkan aplikasi berbasis web yang disebut dengan *Indonesian Documents Similarity Measure.*

Dalam merancang dan mengimplementasikan rancangan aplikasi, digunakan metode *waterfall* atau yang sering disebut dengan *classic life cycle model*. Model *waterfall* ini merupakan model klasik yang bersifat sistematis atau berurutan dalam membangaun perangkat lunak. Model tersebut meliputi beberapa tahapan yakni: (1) *requirements definition*, (2) *system and software design*, (3) *implementation and unit testing* dan (4) *integration and system testing*.

Berdasarkan hasil yang didapat dari pengujian, dapat disimpulkan aplikasi ini telah berhasil menerapkan metode *tokenisation, stopword removal, enhanced confix stemmer* dan model ruang vektor dengan baik. Hasil lain yang didapat adalah aplikasi ini efektif digunakan pada dokumen non-sastra dengan penggunaan kosa kata yang tidak terlalu jauh berbeda.

Kata Kunci : Kesamaan Antar Dokumen, Model Ruang Vektor, *Enhanced Confix Stemmer, Web Based*

ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012

ABSTRACT

The aim of this research is to design and develop a web based application that implemented vector space mode. This research uses tokenisation, stopword removal and enhanced confix stemmer for pre processing. Meanwhile, vector space model are implemented for similarity measure. The result of implementation in this research is a web based application named *Indonesian Documents Similarity Measure*.

In designing and implementing the application design, it was used a waterfall method which is usually called as a classic life cycle model. This is a classic model which creates the software systematically and sequentially that includes some stages namely: (1) requirements definition, (2) system and software design, (3) implementation and unit testing, and (4) integration and testing system.

Based on the results of the testing, this application is successfully implemented tokenisation, stopword removal, enhanced confix stemmer and vector space model methods. The another result from the testing prove this application is effective for non-fiction documents with not so much difference in vocabulary.

Keywords : Documents Similarity, Vector Space Model Enhanced Confix Stemmer, Web Based

1. PENDAHULUAN

Information retrieval merupakan ilmu yang mempelajari bagaimana menemukan informasi yang relevan terhadap kebutuhan dari suatu kumpulan informasi yang berjumlah besar. Informasi atau data yang dicari dapat berupa teks, image, audio, video dan lain-lain Koleksi data teks yang dapat dijadikan sumber pencarian juga dapat berupa pesan teks seperti e-mail, fax dan dokumen berita, bahkan dokumen yang beredar di internet. Jumlah koleksi dokumen yang besar sebagai sumber pencarian menjadi alasan dibutuhkannya suatu sistem yang dapat membantu user menemukan dokumen yang relevan dalam waktu singkat dan tepat.

Pencarian informasi dari data berupa teks (dokumen) merupakan yang paling populer dan paling sering ditemui dari sekian banyak tipe data atau informasi yang dapat ditangani *information retrieval*. Konsep kerja dari *information retrieval* adalah

mencari similaritas antara dokumen dengan dokumen lainnya, atau dokumen dengan kata-kata yang menjadi acuan pencarian (*query*) dari pengguna *information retrieval.* Mempelajari bagaimana mencari nilai kesamaan/similaritas adalah hal yang paling penting untuk meningkatkan performa dari suatu sistem *information retrieval*, baik dari segi efektifas, efisien dan ketepatan hasil.

Perhitungan kesamaan dokumen yang umum digunakan adalah model ruang vektor. Model ruang vektor menghitung kesamaan dengan cara mengukur sudut yang terbentuk antara vektor dokumen dengan vektor *query*. Semakin kecil sudut yang dibentuk suatu vektor dokumen dengan vektor *query*, dokumen tersebut semakin sama dengan *query*, yang berarti dokumen tersebut semakin relevan dengan *query*.

Bahasa menjadi salah satu batasan dalam *information retrieval* yang menangani masalah teks. *Similarity measure* antar dokumen yang menggunakan bahasa yang berbeda tidak akan menghasilkan hasil yang tepat. Batasan itulah yang membuat *information retrieval* hanya efektif untuk dipakai di satu bahasa

Sesuai dengan paparan diatas, penulis tertarik untuk mengembangkan suatu aplikasi untuk mengukur tingkat kesamaan antar dua buah dokumen (*document similarity measure*) menggunakan model ruang vektor. Aplikasi yang dikembangkan nantinya khusus untuk dokumen berbahasa Indonesia.

2. KAJIAN PUSTAKA

2.1 *Documents Similarity Measure*

Document Similarity Measure membandingkan antara kemiripan dengan dua dokumen, jadi kita anggap salah satu dokumen, dimana disini dokumen pertama sebagi query input pengguna. Tetapi untuk perhitungan, dokumen 1 dan 2 tetap dijadikan

ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012

kesatuan dokumen yang akan dibandingkan untuk menghindari proses kesalahan perhitungan matematika. Hanya saja proses pembandingan tidak dilakukan terhadap dokumen pertama yang dijadikan query. Terdapat dua proses utama yang menjadi tahapan Documents Similarity Measure, yaitu *indexing* dan *matching*. *Indexing* adalah proses persiapan yang dilakukan terhadap dokumen sehingga dokumen siap untuk diproses.

2.2 *Indexing*

Tahap-tahap yang terjadi dalam proses *indexing* antara lain :

1. *Tokenisation*

Merupakan pengubahan dokumen mentah menjadi *term* dengan cara menjadi semua kata dalam dokumen tersebut menjadi *lowercase* dan membuang semua tanda baca dan simbol.

2. *Stop-word Removal*

Pembuangan kata penghubung dan kata yang tidak memiliki makna berarti yang ada di dalam dokumen.

3. Stemming

Proses mengubah macam-macam bentukan kata menjadi satu bentuk yang sama. Proses *stemming* sangat bergantung pada bahasa yang dipakai dalam penulisan dokumen, karena semua bahasa memiliki aturan penulisan tersendiri. Algoritma yang dipakai dalam proses *stemming* pada penelitian ini adalah *enchanced confix stemmer* yang merupakan pengembangan dari algoritma *confix stemmer*.

4. Termweighting

Merupakan proses menentukan bobot *term* dalam setiap dokumen. Penentuan bobot ini tergantung dari pemodelan yang digunakan oleh sistem.

ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012

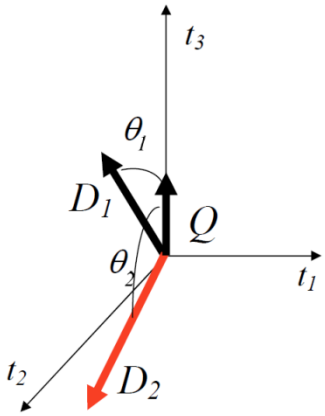
2.3 *Matching*

Matching adalah proses membandingkan dua dokumen dengan term yang didapat dari cara diatas dengan menggunakan metode tertentu. Pada penelitian ini, penulis menggunakan Model Ruang Vektor.

Model ruang vektor pertama kali diperkenalkan oleh Salton dan Lesk (1986) untuk SMART *retrieval system*. Ide yang mendasari model ruang vektor adalah masing-masing *query* dan dokumen dianggap sebagai sebuah vektor n-dimensi. Tiap dimensi pada vektor tersebut diwakili oleh satu *term*. *Term* yang digunakan biasanya berpatokan kepada *term* yang ada pada *query*, sehingga *term* yang ada pada dokumen tetapi tidak ada pada *query* biasanya diabaikan.

Perhitungan kesamaan antara vektor *query* dan vektor dokumen dilihat dari sudut yang paling kecil. Sudut yang dibentuk oleh dua buah vektor dapat dihitung dengan melakukan perkalian dalam (*inner product*), sehingga rumus relevansinya adalah

$$R(Q, D) = \cos\theta = \frac{Q \cdot D}{|Q||D|}$$



Gambar 1 Contoh Tampilan Ruang vektor dari Dokumen dan Query

3. ANALISIS DAN PERANCANGAN

3.1 Analisis Masalah dan Usulan Solusi

Pengukuran tingkat kesamaan dokumen umumnya memiliki permasalahan pada bagian *pre-processing*. Penggunaan metode *stemming* belum sering digunakan,

walaupun digunakan, biasanya masih menggunakan algoritma Nazief. Daftar *stopword (stoplist)* umumnya hanya mencantumkan kata-kata hubung, padahal ada kata lain yang bukan merupakan kata hubung juga tidak memiliki makna yang berarti bagi suatu kalimat. Atas dasar permasalahan tersebut, peneliti menggunakan metode *stemming* dengan algoritma *enhanced confix stemmer* serta penggunaan *stopword* yang tidak saja terdiri dari kata hubung.

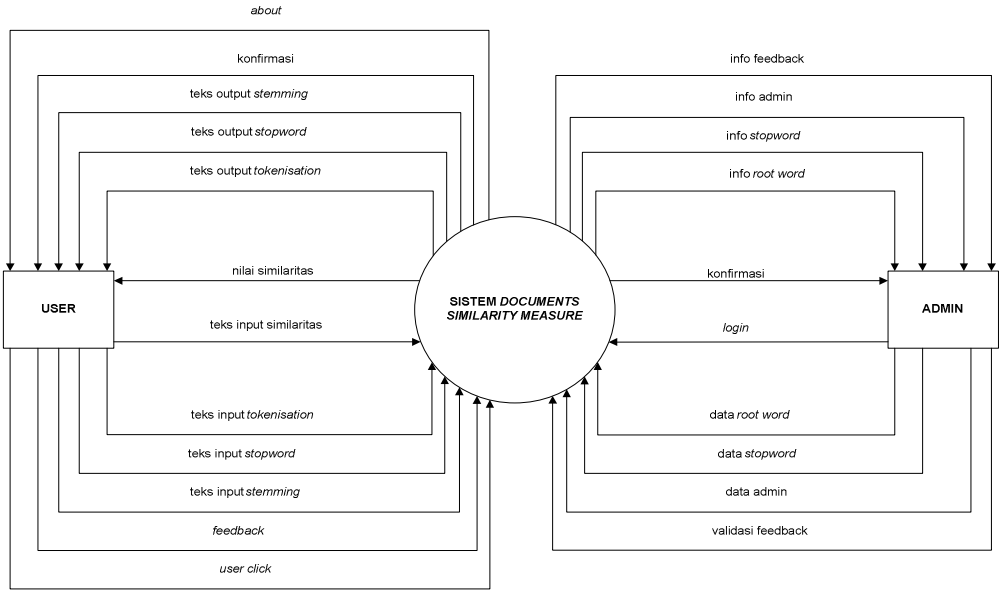
3.2 Analisis Perangkat Lunak

Perangkat lunak yang akan dikembangkan diharapkan dapat menangani proses perhitungan kesamaan dokumen menggunakan model ruang vektor, *pre-processing (tokenisation, stopword removal, stemming), feedback stemming* dan *stopword,* pengelolaan data sistem meliputi data admin, kata dasar dan *stopword* serta proses validasi *feedback.* Tujuan dari pengembangan perangkat lunak ini adalah untuk dapat mengimplementasikan proses-proses tadi. Masukan dari perangkat lunak ini adalah teks untuk proses perhitungan kesamaan dokumen dan *pre-processing*, data admin, data kata dasar, data *stopword, feedback stemming, feedback stopword,* validasi *feedback stemming* dan validasi *feedback stopword.* Keluaran dari perangkat lunak ini adalah nilai dari hasil dari proses perhitungan kesamaan dokumen, teks hasil *pre-processing*, informasi tentang admin, informasi tentang kata dasar, informasi *stopword,* konfirmasi dari *feedback* oleh *user*, informasi validasi *feedback* oleh admin.

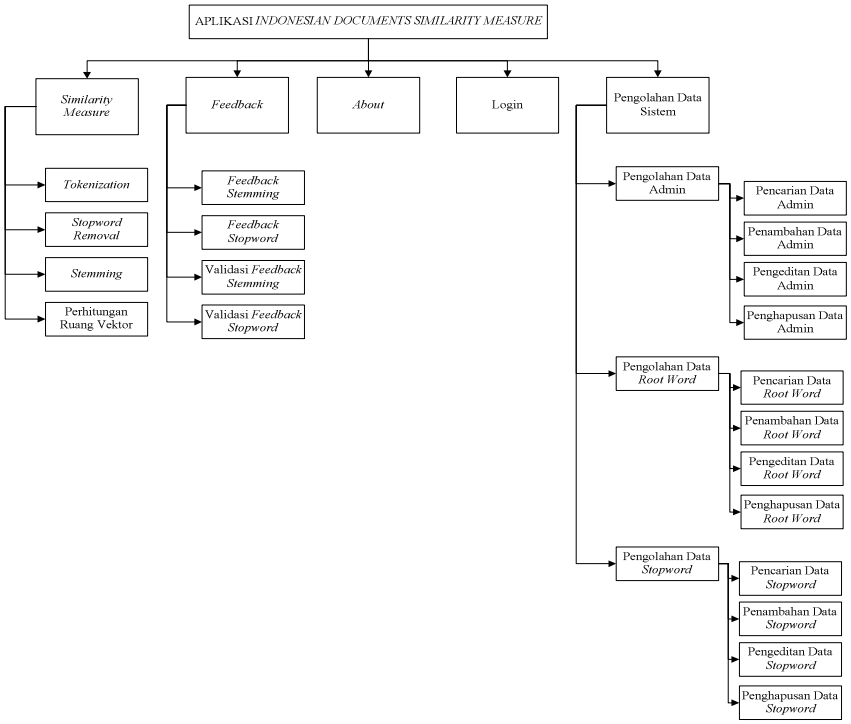
3.3 Perancangan Perangkat Lunak

Perancangan arsitektur perangkat lunak menggambarkan bagian-bagian modul, struktur ketergantungan antar modul dan hubungan antar model dari perangkat lunak yang dibangun. Pada bagian ini terdapat diagram kontes dan *structure chart* sebagai kendali fungsional yang digambarkan seperti Gambar 2 dan Gambar 3 untuk perangkat lunak *Documents Similarity Measure.*

ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012



Gambar 2 Diagram Konteks Perangkat Lunak *Indonesian Documents Similarity Measure*



Gambar 3 Structure Chart Perangkat Lunak *Indonesian Documents Similarity Measure*

ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012

4. IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Perangkat Lunak

Data Flow Diagram (DFD) dan Rancangan Arsitektur Perangkat Lunak *Indonesian Documents Similarity Measure* diimplementasikan menggunakan bahasa pemrograman web. Berikut ini tabel pemetaan unit serta tampilan Halaman Utama dari Perangkat Lunak *Indonesian Documents Similarity Measure.*

Tabel 1 Pemetaan Unit Implementasi

No Unit	Nama Unit	Penjelasan Unit
Unit 1	*about.php*	File PHP ini berfungsi untuk menampilkan halaman *about.*
Unit 2	*echanced_confix_stemmer.php*	File PHP ini menyimpan fungsi untuk melakukan proses *stemming.*
Unit 3	*do_process.php*	File PHP ini berfungsi untuk menampilkan hasil dari masing-masing *pre-processing.*
Unit 4	*dsm.php*	File PHP ini merupakan halaman hasil dari *similarity measure*.
Unit 5	*feedback_stem.php*	File PHP ini merupakan antarmuka user untuk memasukkan *feedback stemming.*
Unit 6	*feedback_stopword.php*	File PHP ini berfungsi filter *feedback stopword* dan memasukkan *feedback* ke database.
Unit 7	*index.php*	File PHP ini merupakan halaman awal *Documents Similarity Measure* pada sesi user.
Unit 8	*layout.php*	File PHP ini menyimpan fungsi yang mendefinisikan bagian dari layout web untuk sesi user.
Unit 9	*lib_function.php*	File PHP ini menyimpan fungsi-fungsi utama dari aplikasi *documents similarity measure.*
Unit 10	*process.php*	File PHP ini merupakan halaman input dari *pre-processing.*
Unit 11	*save_rt.php*	File PHP ini berfungsi filter *feedback stemming* dan memasukkan *feedback* ke database.
Unit 12	*similarity_measure.php*	File PHP ini berfungsi filter *feedback stemming* dan memasukkan *feedback* ke database.
Unit 13	*style.css*	File CSS ini mendefinisikan tampilan dari web pada bagian user.
Unit 14	*add.php*	File PHP ini berfungsi untuk menangani proses penambahan data.
Unit 15	*add_admin_form.php*	File PHP ini berfungsi untuk menampilkan form tambah data admin.
Unit 16	*add_root_word.php*	File PHP ini berfungsi untuk menampilkan form tambah data *rootword*.

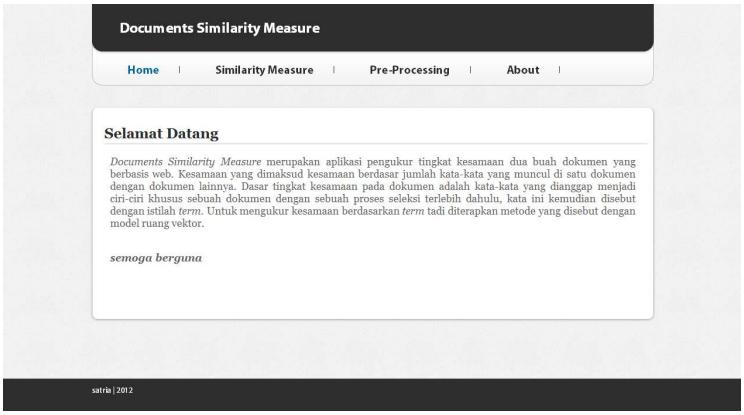
ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012

No Unit	Nama Unit	Penjelasan Unit
Unit 17	*add_stopword.php*	File PHP ini berfungsi untuk menampilkan form tambah data *stopword*.
Unit 18	*admin.css*	File CSS ini mendefinisikan tampilan pada sesi login admin.
Unit 19	*admin_layout.php*	File PHP ini menyimpan fungsi yang mendefinisikan bagian dari layout web untuk sesi admin.
Unit 20	*delete.php*	File PHP ini berfungsi untuk menangani proses penghapusan data.
Unit 21	*edit.php*	File PHP ini berfungsi untuk menangani proses pengeditan data.
Unit 22	*edit_admin_form.php*	File PHP ini berfungsi untuk menampilkan form edit data admin.
Unit 23	*edit_root_word_form.php*	File PHP ini berfungsi untuk menampilkan form edit data *root word*.
Unit 24	*edit_stopword.php*	File PHP ini berfungsi untuk menampilkan form edit data *stopword*.
Unit 25	*feed_stem_validation.php*	File PHP ini berfungsi untuk menangani proses validasi *feedback stemming.*
Unit 26	*feed_stop_validation.php*	File PHP ini berfungsi untuk menangani proses validasi *feedback stopword.*
Unit 27	*form_add.php*	File PHP ini berfungsi sebagai template untuk proses tambah data sistem.
Unit 28	*form_edit.php*	File PHP ini berfungsi sebagai template untuk proses edit data sistem.
Unit 29	*home_page.php*	File PHP ini merupakan halaman yang tampil jika admin berhasil melakukan login.
Unit 30	*index.php*	File PHP ini merupakan antarmuka/validasi admin untuk masuk ke sistem administator.
Unit 31	*logout.php*	File PHP ini berfungsi untuk melakukan logout dari halaman admin.
Unit 32	*search.php*	File PHP ini berfungsi untuk melakukan pencarian semua data sistem.
Unit 33	*style.css*	File CSS ini berfungsi untuk mendefinisikan tampilan halaman admin.
Unit 34	*view.php*	File PHP ini berfungsi sebagai template untuk view tiap data sistem.
Unit 35	*view_admin.php*	File PHP ini berfungsi untuk menampilkan data admin.
Unit 36	*view_feedback_stem.php*	File PHP ini berfungsi untuk menampilkan data *feedback stemming*.
Unit 37	*view_feedback_stop.php*	File PHP ini berfungsi untuk menampilkan data *feedback stopword*.

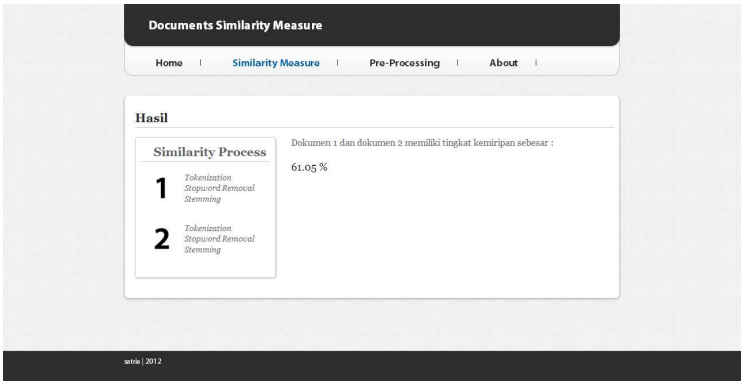
ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012

No Unit	Nama Unit	Penjelasan Unit
Unit 38	*view_root_word.php*	File PHP ini berfungsi untuk menampilkan data *root word*.
Unit 39	*view_stopword.php*	File PHP ini berfungsi untuk menampilkan data *stopword*.

Gambar 4, 5 dan 5 dibawah ini merupakan *screenshoot* dari tampilan aplikasi *documents similarity measure* pada sesi user dan admin.

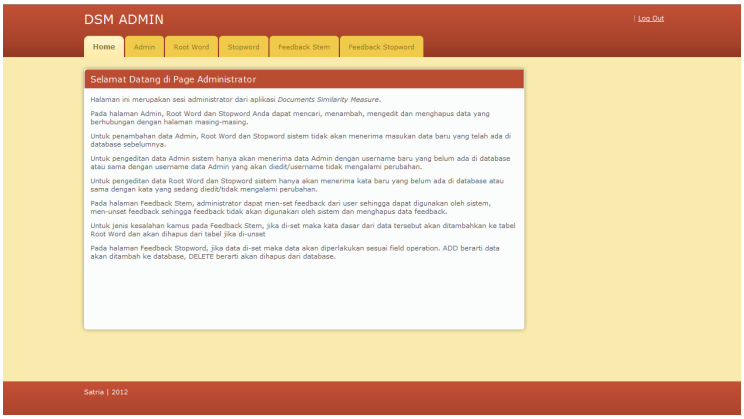


Gambar 4 Implementasi Home Page *DSM*



Gambar 5 Implementasi Halaman *Similarity Measure*

ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012



Gambar 6 Implementasi Halaman Admin

4.2 Pengujian Perangkat Lunak

Pengujian *similarity measure* menggunakan 6 buah file dokumen, dengan 2 topik bahasan yang berbeda. 3 buah file dokumen dengan tema teknologi IT dan 3 buah file dengan tema kesehatan. Berikut merupakan kilasan tentang keenam buah dokumen tersebut.

a) Dokumen pertama (A) tentang topik Pengertian PHP dengan kata sebanyak 419 buah.
b) Dokumen kedua (B) tentang topik Pengertian PHP dengan kata sebanyak 332 buah.
c) Dokumen ketiga (C) tentang topik Pengertian PHP dengan kata sebanyak 518 buah.
d) Dokumen keempat (D) tentang topik Dampak Merokok dengan kata sebanyak 634 buah.
e) Dokumen kelima (E) tentang topik Dampak Merokok dengan kata sebanyak 561 buah.
f) Dokumen keenam (F) tentang topik Dampak Merokok dengan kata sebanyak 763 buah.

Selain pengujian antar dokumen-dokumen diatas, juga dilakukan perhitungan kesamaan dokumen pertama dengan dokumen hasil campuran semua dokumen kedua dan setengah dari konten dokumen pertama. Setengah konten dari dokumen didapat

dengan cara meng*copy* setengah jumlah kata pertama dari keseluruhan jumlah kata dalam dokumen. Dibawah ini merupakan tabel hasil pengujian kemiripan antar dokumen tersebut serta kemungkinan kombinasinya.

Tabel 2 Hasil Pengujian Proses *Similarity Measure*

No	Dokumen 1	Dokumen 2	Hasil Similaritas
1	Dokumen A	Dokumen A	100 %
2	Dokumen A	Dokumen B	61,05 %
3	Dokumen A	Dokumen C	14,21 %
4	Dokumen A	Dokumen D	1,28 %
5	Dokumen A	Dokumen E	1,31 %
6	Dokumen A	Dokumen F	2,74 %
7	Dokumen B	Dokumen B	100 %
8	Dokumen B	Dokumen C	13,8 %
9	Dokumen B	Dokumen D	2,34 %
10	Dokumen B	Dokumen E	1,58 %
11	Dokumen B	Dokumen F	2,93 %
12	Dokumen C	Dokumen C	100 %
13	Dokumen C	Dokumen D	2,2 %
14	Dokumen C	Dokumen E	2,06 %
15	Dokumen C	Dokumen F	3,27 %
16	Dokumen D	Dokumen D	100 %
17	Dokumen D	Dokumen E	16,46 %
18	Dokumen D	Dokumen F	16,52 %
19	Dokumen E	Dokumen E	100 %
20	Dokumen E	Dokumen F	16,45 %
21	Dokumen F	Dokumen F	100 %
22	Dokumen A	Dokumen A + 50% konten dari Dokumen A	100 %
23	Dokumen A	Dokumen B + 50% konten dari Dokumen A	84,81 %
24	Dokumen A	Dokumen C + 50% konten dari Dokumen A	35,7 %
25	Dokumen A	Dokumen D + 50% konten dari Dokumen A	27,98 %
26	Dokumen A	Dokumen E + 50% konten dari Dokumen A	24,05 %
27	Dokumen A	Dokumen F + 50% konten dari Dokumen A	27,6 %
28	Dokumen B	Dokumen B + 50% konten dari Dokumen B	100 %
29	Dokumen B	Dokumen C + 50% konten dari Dokumen B	27,53 %

ISSN 2252-9063
Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI)
Volume 1, Nomor 3, Juli 2012

No	Dokumen 1	Dokumen 2	Hasil Similaritas
30	Dokumen B	Dokumen D + 50% konten dari Dokumen B	20,13 %
31	Dokumen B	Dokumen E + 50% konten dari Dokumen B	16,72 %
32	Dokumen B	Dokumen F + 50% konten dari Dokumen B	19,48 %
33	Dokumen C	Dokumen C + 50% konten dari Dokumen C	100 %
34	Dokumen C	Dokumen D + 50% konten dari Dokumen C	26,04 %
35	Dokumen C	Dokumen E + 50% konten dari Dokumen C	22,66 %
36	Dokumen C	Dokumen F + 50% konten dari Dokumen C	25,98 %
37	Dokumen D	Dokumen D + 50% konten dari Dokumen D	100 %
38	Dokumen D	Dokumen E + 50% konten dari Dokumen D	27,9 %
39	Dokumen D	Dokumen F + 50% konten dari Dokumen D	32,39 %
40	Dokumen E	Dokumen E + 50% konten dari Dokumen E	100 %
41	Dokumen E	Dokumen F + 50% konten dari Dokumen E	46,69 %
42	Dokumen F	Dokumen F + 50% konten dari Dokumen F	100 %

Berdasarkan hasil diatas, dapat ditarik kesimpulan bahwa nilai similaritas yang relatif besar didapatkan dari dokumen dengan kesamaan kosa kata dalam jumlah yang relatif banyak. Hal lain yang didapatkan adalah, aplikasi ini tidak menghiraukan kemunculan *term*, seperti terlihat pada hasil uji no 22, 28, 33, 37, 40 dan 42. Hasil similaritas yang didapat sama dengan sebelum dokumen dimodifikasi kontennya. Peningkatan nilai similaritas terjadi pada pengujian dokumen dengan konten yang telah dimodifikasi, sebesar 20-30%. Hal ini sesuai dengan konsep perhitungan model ruang vektor.

5. PENUTUP

5.1 Simpulan

Berdasarkan hasil analisis, implementasi dan pengujian pada penelitian ini, maka dapat diambil simpulan bahwa aplikasi *Documents Similarity Measure* yang dikembangkan berbasis web ini dengan menggunakan metode model ruang vektor, *tokenisation, stopword removal, stemming* telah mampu menerapkan metode yang digunakan dengan baik serta mampu mengukur tingkat kesamaan dokumen dengan hasil yang cukup akurat sesuai dengan konsep model ruang vektor.

5.2 Saran

Berdasarkan pengamatan penulis, disarankan bagi pembaca yang ingin mengembangkan aplikasi ini agar dapat menggunakan metode lain yang dapat mengukur tingkat kesamaan antar dokumen dari segi linguistik, mengembangkan algoritma *stemming* yang digunakan di penelitian ini agar dapat menangani kata bersisipan dan menggunakan tipe file masukan yang lebih beragam.

6. DAFTAR PUSTAKA

Adhi Kerta Mahendra, I Putu. 2008. *Penggunaan Algoritma Semut dan Confix Stripping Stemmer untuk Klasifikasi Dokumen Berita Berbahasa Indonesia.* Jurusan Teknik Informatika. Institut Teknologi Sepuluh November.

Asian, Jelita., Williams, Hugh E., Tahaghohi, S.M.M. 2007. *Stemming Indonesian*. School of Computer Science and Information Technology. RMIT University.

Document Indexing Tutorial. http://www.miislita.com/information-retrieval-tutorial/indexing.html (diakses pada tanggal 3 Desember 2011)

Ivan Widyarsa, Michael. 2007. *Peningkatan Performansi Information Retrieval dengan Menerapkan Relevance Feedback Berbasis Algoritma Genetika.* Department of Informatics. Institut Teknologi Bandung.

Rachmansyah. 2008. *Temu Kembali Informasi (Information Retrieval).* http://rachmansyah.web.id (diakses pada tanggal 8 Desember 2011)

Ramadhany, Taufik. 2008. *Implementasi Kombinasi Model Ruang Vektor dan Model Probabilistik Pada Sistem Temu Balik Informasi.* Department of Informatics. Institut Teknologi Bandung.