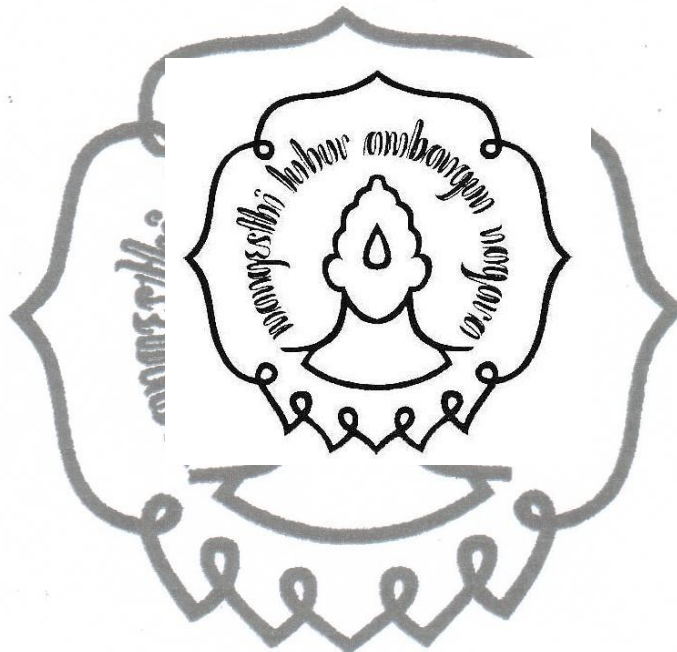


**Analisis Perbandingan *Vector Space Model* dan
Weighted Tree Similarity pada Pencarian Informasi
Ebook Pedoman Pengobatan Dasar di Puskesmas**

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Syarat Mencapai Gelar Strata Satu

Jurusan Informatika



Disusun oleh :

Viko Basmalah Wicaksono

NIM. M0510044

**JURUSAN INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SEBELAS MARET
SURAKARTA**

comm2015 user

SKRIPSI
Analisis Perbandingan *Vector Space Model* dan
***Weighted Tree Similarity* pada Pencarian Informasi**
***Ebook* Pedoman Pengobatan Dasar di Puskesmas**

Disusun oleh :
Viko Basmalah Wicaksono
M0510044

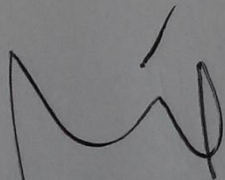
Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji,
Pada tanggal : 21 Januari 2015

PERSETUJUAN PEMBIMBING

Tugas Akhir ini telah disetujui oleh :

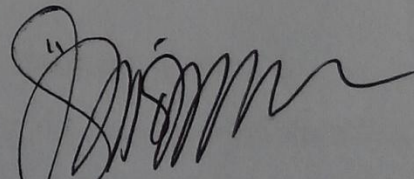
Pembimbing 1,

Pembimbing 2,



Ristu Saptono S.Si, M.T

NIP. 19790210 20021 2 1001



Sari Widya Sihwi, S.Kom., M.T.I

NIP. 19830412 20091 2 2003.

SKRIPSI

**Analisis Perbandingan *Vector Space Model* dan
Weighted Tree Similarity pada Pencarian Informasi
Ebook Pedoman Pengobatan Dasar di Puskesmas**

Disusun oleh :

Viko Basmalah Wicaksono

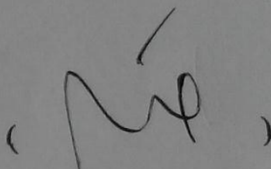
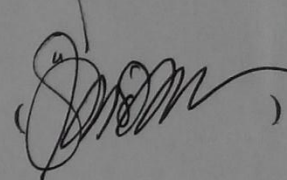
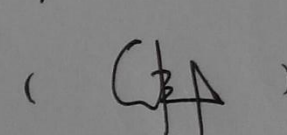
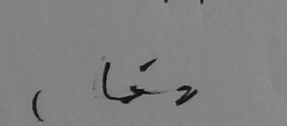
M0510044

Telah dipertahankan di hadapan Dewan Penguji

Pada tanggal : 21 Januari 2015

Susunan Dewan Penguji

1. Ristu Saptono S.Si, M.T
NIP. 19790210 200212 1 001
2. Sari Widya Sihwi, S.Kom., M.T.I
NIP. 19830412 200912 2 003
3. Meiyanto Eko Sulistyo, S.T., M.Eng
NIP. 19770513 200912 1 004
4. Dr. Wiranto, M.Kom., M.Cs.
NIP. 19661230 199302 1 001

()
()
()
()

Disahkan oleh:

Ketua Jurusan Informatika,



Drs. Bambang Harjito M.App.Sc., Ph.D

NIP. 19700217 199702 2 001

MOTTO

*“Jadilah yang teratas, tapi bukan dengan menginjak kepala orang lain
Jadilah yang tertinggi, tapi tidak dengan mencuri tangga orang lain”*

(Si Juki)

“Ilmu tanpa praktek lumpuh, teori tanpa praktek buta”

(Albert Einstein)



commit to user

PERSEMBAHAN

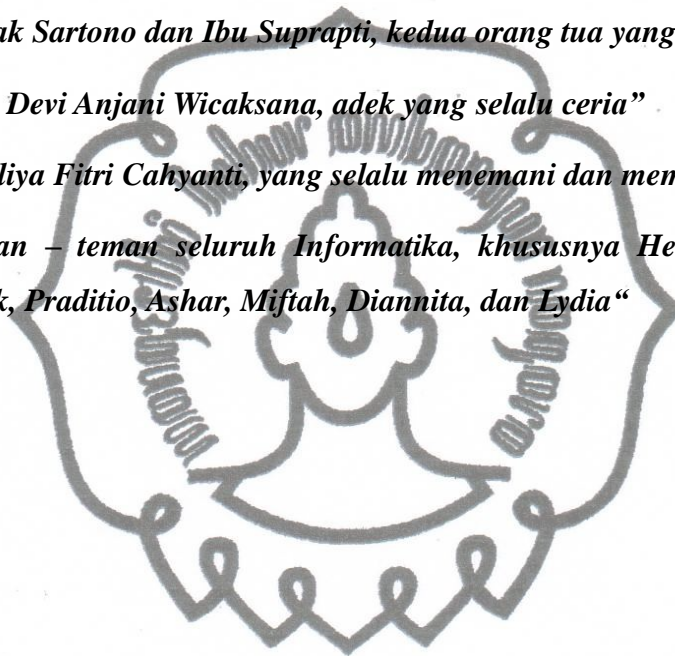
Karya ini Penulis persembahkan kepada:

“Bapak Sartono dan Ibu Suprapti, kedua orang tua yang tercinta”

“Alya Devi Anjani Wicaksana, adek yang selalu ceria”

“Apriliya Fitri Cahyanti, yang selalu menemani dan memberi dukungan”

“Teman – teman seluruh Informatika, khususnya Hedik, Aji, Cerren, Taufik, Praditio, Ashar, Miftah, Diannita, dan Lydia”



commit to user

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Allah SWT, yang hanya karena rahmat dan karunia-Nya, penulis dapat menyelesaikan Tugas Akhir dengan judul “Analisis Perbandingan *Vector Space Model* dan *Weighted Tree Similarity* dengan *Cosine Similarity* pada kasus Pencarian Informasi Pedoman Pengobatan Dasar di Puskesmas ”. Penulis menyadari akan keterbatasan yang dimiliki. Begitu banyak bantuan dan bimbingan yang diberikan dalam penyusunan Tugas Akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada :

1. Ayah, ibu, dan adik yang senantiasa selalu memberikan motivasi, doa dan dukungannya.
2. Bapak Ristu Saptono, S.Si., M.T. selaku Dosen Pembimbing I yang telah memberikan pengarahan selama proses penyusunan Tugas Akhir ini,
3. Ibu Sari Widya Sihwi, S.Kom., M.T.I. selaku Dosen Pembimbing II yang telah memberikan masukan, kritik dan saran yang membangun,
4. Bapak Ibu dosen informatika FMIPA UNS,
5. Bapak dr. Burhanuddin Ichsan, M. Med. Ed, M.Kes selaku ahli pakar yang telah membantu dalam penyelesaian penelitian,
6. Apriliya Fitri Cahyanti yang selalu memberikan doa dan dukungannya,
7. Seluruh teman – teman informatika 2010 yang telah memberikan kebersamaannya dan membantu selama ini,
8. Semua pihak yang tidak dapat disebutkan satu persatu.

Semoga Tugas Akhir ini bermanfaat dan memberikan inspirasi bagi semua pihak yang berkepentingan.

Surakarta, 13 Januari 2015

Penulis

commit to user

***Comparative Analysis of Vector Space Model and
Weighted Tree Similarity in Information Searching of Ebook Basic
Treatment Guidelines at Community Health Center***

Viko Basmalah Wicaksono
Informatics Department. Faculty of Mathematics and Natural Science.
Sebelas Maret University

ABSTRACT

Searching system is one solution that help in getting the wanted information. By searching system, searching process will be more efficient. Searching system on the ebook of basic treatment guidelines at the Community Health Center is very helpfull because there are a lot of disease data in that ebook. To develop a searching system of basic treatment information guideline at the community health center can utilize Vector Space Model (VSM) methods or Weighted Tree Similarity (WTS). This study compared the Vector Space Model (VSM) method and Weighted Tree Similarity (WTS) in searching of basic treatment information guideline. Furthermore, in this study Hamming Distance algorithm is added to determine the effect of system execution time.

This study shows that the WTS is better than the VSM. It can be seen on the test results, value of the WTS precision is better than VSM. Although the value recall of WTS is lower than VSM, it not influently because the more effective searching method that gives the best accuracy value(precision). In testing system, VSM shows average results precision is 44.82983% and recall is 99.08165%. While the WTS value average precision is 52.17332% and recall is 98.61761%. Then in the expert testing, VSM shows average results precision is 46.675% and a recall is 73.6111%. While the VSM precision value is 33.6737% and he recall is 86.8056%.

The effect of using Hamming Distance algorithm is very helpful in speeding up the execution of the system. The average time testing result of VSM with additional Hamming Distance algorithm is 4.512 seconds, where without Hamming Distance is 9.185 seconds. Then on the results of WTS method with Hamming Distance is 6.042 seconds, while without Hamming Distance is 14.421 seconds.

Keywords: *Basic Treatment Guidelines at Community Health Center, Searching system, Vector Space Model, Weighted Tree Similarity*

**Analisis Perbandingan *Vector Space Model* dan
Weighted Tree Similarity pada Pencarian Informasi
Ebook Pedoman Pengobatan Dasar di Puskesmas**

**Viko Basmalah Wicaksono
Jurusan Informatika. Fakultas Matematika dan Ilmu Pengetahuan Alam.
Universitas Sebelas Maret**

ABSTRAK

Sistem pencarian merupakan salah satu solusi yang dapat membantu dalam mendapatkan informasi yang diinginkan. Dengan sistem pencarian, proses pencarian informasi akan menjadi lebih efisien. Sistem pencarian informasi pada *ebook* pedoman pengobatan di puskesmas sangat dibutuhkan karena terdapat banyak data penyakit di dalamnya. Dalam mengembangkan sistem pencarian pada pedoman pengobatan di puskesmas, dapat memanfaatkan metode *Vector Space Model* (VSM) atau *Weighted Tree Similarity* (WTS). Penelitian ini membandingkan metode VSM dengan WTS untuk mendapatkan metode terbaik. Selain itu, ditambahkan algoritma *Hamming Distance* untuk mengetahui pengaruh eksekusi waktu sistem.

Penelitian ini menunjukkan bahwa WTS lebih baik dibandingkan VSM. Hal ini dapat dilihat pada hasil pengujian, nilai *precision* pada WTS lebih baik dibandingkan VSM. Meskipun nilai *recall* pada WTS lebih rendah, hal ini tidak berpengaruh karena pada hasil pencarian yang efektif adalah yang memberikan nilai ketepatan (*precision*) terbaik. Pada pengujian sistem, VSM menunjukkan hasil nilai rata – rata *precision* sebesar 44,82983 % dan *recall* sebesar 99,08165 %. Sedangkan pada WTS nilai rata – rata *precision* sebesar 52,17332% dan *recall* sebesar 98,61761%. Kemudian pada pengujian pakar menunjukkan *precision* WTS dengan rata – rata sebesar 46,675% dan *recall* sebesar 73,6111%. Sedangkan nilai *precision* VSM sebesar 33,6737% dan nilai *recall* sebesar 86,8056%.

Algoritma *Hamming Distance* sangat membantu dalam mempercepat eksekusi sistem. Pengaruh penggunaan algoritma *Hamming Distance* pada VSM memberikan hasil dengan rata – rata waktu pengujian adalah 4,512 detik, sedangkan tanpa *Hamming Distance* adalah 9,185 detik. Kemudian pada hasil pengujian waktu metode WTS dengan *Hamming Distance* memberikan hasil rata – rata adalah 6,042 detik, sedangkan tanpa *Hamming Distance* adalah 14,421 detik.

Kata kunci : *Pedoman Pengobatan Dasar Puskesmas, Sistem pencarian, Vector Space Model, Weighted Tree Similarity*

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
MOTTO	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR.....	vi
ABSTRACT.....	vii
ABSTRAK.....	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Dasar Teori	4
2.1.1. <i>Text Preprocessing</i>	4
2.1.2. <i>Cosine Similarity</i>	7
2.1.3. <i>Vector Space Model</i>	7
2.1.4. <i>Weighted Tree Similarity</i>	9
2.1.5. <i>Hamming Distance</i>	10
2.1.6. <i>Precision dan Recall</i>	10
2.1.7. Pedoman Pengobatan Dasar di Puskesmas	11
2.2 Penelitian Terkait.....	12
2.3 Kerangka Pemikiran	14
BAB III METODOLOGI PENELITIAN	16
3.1. Studi Literatur.....	16

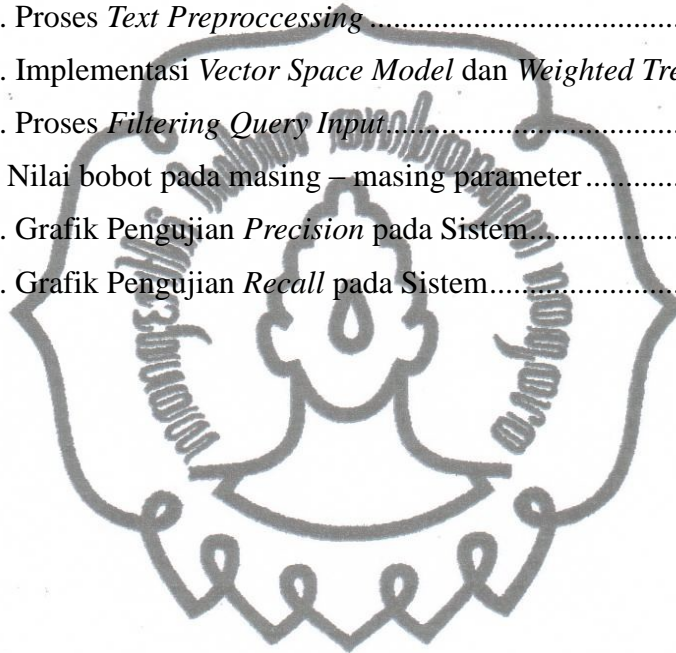
3.2.	Pengumpulan Data	16
3.3.	Penerapan Metode	16
3.3.1.	<i>Preprocessing Query Input</i>	17
3.3.2.	Implementasi metode VSM dan <i>Weighted Tree Similarity</i>	17
3.3.2.1.	Implementasi <i>Hamming Distance</i>	18
3.3.2.2.	Implementasi <i>Vector Space Model</i>	18
3.3.2.3.	Implementasi <i>Weighted Tree Similarity</i>	19
3.4.	Pengembangan Aplikasi	20
3.5.	Hasil dan Pembahasan.....	20
3.5.1.	Pengujian Sistem	20
3.5.2.	Pengujian dengan Pakar	21
BAB IV HASIL DAN PEMBAHASAN		22
4.1.	Deskripsi Data	22
4.2.	Penerapan Metode	22
4.2.1.	<i>Preprocessing Query Input</i>	22
4.2.2.	Indexing Data penyakit	23
4.2.3.	Implementasi Metode VSM dan <i>Weighted Tree Similarity</i>	25
4.2.2.1.	Penerapan Algoritma <i>Hamming Distance</i>	25
4.2.2.2.	Implementasi <i>Vector Space Model</i>	26
4.2.2.3.	Implementasi <i>Weighted Tree Similarity</i>	27
4.3.	Pengembangan Aplikasi	32
4.4.	Hasil dan Pembahasan.....	32
4.4.1.	Pengujian Sistem	32
4.4.1.1.	Pembentukan Jumlah <i>Keyword</i>	32
4.4.1.2.	Penentuan <i>Threshold</i>	33
4.4.1.3.	Analisis Hasil Pengujian Sistem.....	33
4.4.2.	Analisis Pengujian dengan Pakar	36
4.4.3.	Analisis Metode Terbaik.....	38
BAB V PENUTUP.....		39
5.1.	Kesimpulan.....	39
5.2.	Saran.....	39
DAFTAR PUSTAKA		40

DAFTAR TABEL

Tabel 2.1. Kombinasi Awalan Akhiran yang tidak Dijiinkan.....	6
Tabel 2.2. Cara Menentukan Tipe Awalan Untuk Kata Yang Diawali “te-”	6
Tabel 2.3. Jenis Awalan Berdasarkan Tipe Awalannya	6
Tabel 2.4. Keterkaitan penelitian dengan penelitian sebelumnya	14
Tabel 3.1. Proses dan Metode pada Vector Space Model	19
Tabel 4.1. Penghapusan Format dan Markup	22
Tabel 4.2. Stemming	23
Tabel 4.3. Penghitungan Frekuensi <i>Query Input</i>	23
Tabel 4.4. Contoh data asli yang belum diindex	23
Tabel 4.5 Contoh Dokumen yang sudah di-index di <i>Database</i>	24
Tabel 4.6. Contoh penerapan algoritma <i>Hamming distance</i>	26
Tabel 4.7. Contoh Pembobotan VSM pada <i>query input</i> dengan data penyakit.....	26
Tabel 4.8. Contoh pembobotan TF pada <i>Query</i> dan parameter nama penyakit....	28
Tabel 4.9. Contoh pembobotan TF pada <i>Query</i> dan parameter definisi	28
Tabel 4.10. Contoh pembobotan TF pada <i>Query</i> dan parameter penyebab	28
Tabel 4.11. Contoh pembobotan TF pada <i>Query</i> dan parameter gambaran klinis	29
Tabel 4.12. Contoh pembobotan TF pada <i>Query</i> dan parameter diagnosis	29
Tabel 4.13. Jumlah Percobaan Sistem.....	32
Tabel 4.14. Hasil Pembentukan <i>Threshold</i>	33
Tabel 4.15. Hasil pengujian <i>Precision</i> dan <i>Recall</i> pada VSM	33
Tabel 4.16. Hasil <i>Precision</i> dan <i>Recall</i> pada <i>Weighted Tree Similarity</i>	34
Tabel 4.17. Hasil Pengujian Waktu dengan metode VSM	35
Tabel 4.18. Hasil Pengujian Waktu dengan metode <i>Weighted Tree Similarity</i>	36
Tabel 4.19. Hasil Pengujian pakar	36

DAFTAR GAMBAR

Gambar 2.1. Contoh representasi dokumen tiga dimensi pada VSM	8
Gambar 2.2. Contoh Reprerstasi VSM pada matriks term-document.....	8
Gambar 2.3. Contoh perhitungan dasar kemiripan <i>tree</i> (Perdana, 2014).....	10
Gambar 2.4. Diagram Penjelasan <i>Recall</i> (Clarke & Willett, 1997)	11
Gambar 2.5. Diagram penjelasan <i>Precision</i> (Clarke & Willett, 1997)	11
Gambar 3.1. Proses Penerapan Metode pada Sistem.....	16
Gambar 3.2. Proses <i>Text Preprocessing</i>	17
Gambar 3.3. Implementasi <i>Vector Space Model</i> dan <i>Weighted Tree Similarity</i>	18
Gambar 4.1. Proses <i>Filtering Query Input</i>	23
Gambar 4.2 Nilai bobot pada masing – masing parameter	31
Gambar 4.3. Grafik Pengujian <i>Precision</i> pada Sistem.....	34
Gambar 4.4. Grafik Pengujian <i>Recall</i> pada Sistem.....	35



DAFTAR LAMPIRAN

LAMPIRAN A. TAMPILAN <i>USER INTERFACE</i> SISTEM	42
LAMPIRAN B. HASIL PENGUJIAN SISTEM	466
LAMPIRAN C. HASIL PENGUJIAN PAKAR	666
LAMPIRAN D. HASIL REKAPITULASI PENGUJIAN PAKAR	788



commit to user