IDENTIFIKASI KECOCOKAN DOKUMEN TANYA JAWAB MENGGUNAKAN METODE *TEXT MINING* DAN *VECTOR SPACE MODEL.*

SKRIPSI

Oleh:

Masti Fatchiyah Maharani
1434010077

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"

JAWA TIMUR

2018

LEMBAR PENGESAHAN
SKRIPSI

Judul : IDENTIFIKASI KECOCOKAN DOKUMEN TANYA JAWAB MENGGUNAKAN METODE *TEXT MINING* DAN *VECTOR SPACE MODEL*

Oleh : MASTI FATCHIYAH MAHARANI

NPM : 1434010077

Telah Diseminarkan Dalam Ujian SKripsi Pada:
Hari Senin, Tanggal 21 Mei 2018

Menyetujui

Dosen Pembimbing

1.

Fetty Tri Anggraeny., S.Kom, M.Kom
NPT : 3 82020 6020 81

2.

Intan Yuniar P., S.Kom, M.Sc
NPT : 3 8006 04019 81

Dosen Penguji

1.

Budi Nugroho., S.Kom, M.Kom
NPT : 3 8009 050 205 1

2.

Eva Yulia P.S.Kom, M.Kom
NPT : 3 8907 130 346 1

3.

Mohammad Idhom S.P, S.Kom,MT
NPT : 3 8303 100 285 1

Mengetahui :

Dekan
Fakultas Ilmu Komputer,

Dr. Ir. Ni Ketut Sari, MT
NIP. 19650731 199203 2001

Koordinator Program Studi
Teknik Informatika,

Budi Nugroho, S.Kom, M.Kom
NPT. 380090502051

IDENTIFIKASI KECOCOKAN DOKUMEN TANYA JAWAB MENGGUNAKAN METODE TEXT MINING DAN ALGORITMA VECTOR SPACE MODEL (VSM).

Nama Mahasiswa	**: Masti Fatchiyah Maharani**
NPM	**: 1434010077**
Program Studi	**: Teknik Informatika**
Dosen Pembimbing	**: Fetty Tri Anggaraeny S.Kom, M.Kom**
	Intan Yuniar P. Skom, M.sc

Abstrak

Information Retrieval (IR) atau sistem temu kembali merupakan sebuah sistem yang bekerja untuk menemukan informasi yang relevan dengan apa yang dicari oleh pengguna. Dalam penelitian yang dilakukan oleh penulis, proses *IR* diterapkan dengan menggunakan metode *text mining* dan algoritma *vector space model*. *Text mining* merupakan proses awal dari alur kerja sebuah sistem temu kembali, yang mana proses ini bertujuan untuk menghasilkan kumpulan kata-kata yang berhubungan dengan dokumen yang dicari. Proses *IR* kemudian dilanjutkan dengan proses mencari nilai kemiripan antar dokumen yang relevan menggunakan algoritma *vector space model.* Algoritma *vector space model* merupakan sebuah algoritma yang menggambarkan sebuah dokumen sebagai sebuah bentuk *vector* yang memiliki jarak dan arah.

Dalam penerapannya, algoritma ini melalui beberapa tahapan seperti *indexing* dokumen yang dilakukan dalam proses *text mining*, pembobotan term yang dalam penelitian ini dilakukan dengan menggunakan algoritma TF-IDF, dan perhitungan kesamaan. Ada beberapa jenis metode yang digunakan dalam menghitung kesamaan dokumen dalam algoritma *vector space model*, salah satunya adalah *cosine similarity* yang digunakan dalam penelitian ini. *Cosine similarity* mengukur kesamaan antar dua vektor dengan mengambil nilai kosinus dari sudut antara vektor kata kunci dan vektor tiap dokumen.

Dari hasil penelitian ini, diketahui bahwa sistem mampu menghasilkan nilai rata-rata *precision* sebesar 3.04%

Kata kunci: *Information Retrieval (IR), Text Mining, Vector Space Model(VSM).*

KATA PENGANTAR

Segala puji bagi Allah SWT yang Maha pengasih lagi Maha penyayang sebab karena karunia. kebaikan, dan kemudahanNya lah penulis dapat menyelesaikan skripsi yang berjudul "Identifikasi Kecocokan Dokumen Tanya Jawab Menggunakan Metode *Text Mining* dan *Vector Space Model*" sesuai dengan rentang waktu yang direncanakan.

Skripsi ini dibuat sebagai salah satu persyaratan untuk memenuhi salah satu syarat dalam menempuh ujian sarjana pada Fakultas Ilmu Komputer, Program Studi Teknik Informatika di Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Penulis berharap bahwa dengan penyusunan skripsi ini mampu membuka "gerbang" ilmu baru dan memberikan manfaat bagi teman-teman pembaca. Namun, penulis juga menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kesalahan maka dari itu penulis berharap adanya saran dan kritik membangun bagi penulis dan juga semoga penelitian ini dapat dikembangkan dan disempurnakan kedepannya.

Surabaya, Mei 2018

Penulis

UCAPAN TERIMA KASIH

Dalam penyusunan dan penulisan skripsi ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan ini penulis dengan senang hati menyampaikan terima kasih kepada yang terhormat:

1. Bapak Prof. Dr. Ir. H. Teguh Soedarto, M.P selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur.
2. Ibu Dr. Ir. Ni Ketut Sari, M.T. selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Budi Nugroho, S.kom., M.kom., selaku Ketua Progdi Teknik Informatika.
4. Ibu Fetty Tri Anggraeny, S.kom., M.kom., selaku pembimbing I dan Ibu Intan Yuniar Purbasari, S.Kom, M.Sc., selaku pembimbing II, yang dengan sabar, dan ikhlas meluangkan waktu, tenaga dan pikiran memberikan bimbingan, motivasi, arahan, dan saran- yang sangat berharga kepada penulis selama menyusun skripsi.
5. Staff Dosen Teknik Informatika UPN "Veteran" Jawa Timur yang telah membekali penulis dengan berbagai ilmu selama mengikuti perkuliahan sampai akhir penulisan skripsi
6. (Alm) Ayah, Bunda, Abang, dan Mbak yang selalu memberikan dukungan secara moril dan materiil dalam penyusunan skripsi ini
7. Partner skripsi penulis yang telah membantu banyak dalam penyusunan konsep dan pengerjaan skripsi, Khoritul Arif Yanto.

8. Teman-teman yang bersedia mendengarkan keluh kesah serta menghibur penulis dikala jenuh yaitu Widya Sekar Arum dan David Risdianto
9. Teman-teman Teknik Informatika angkatan 2014, khususnya kelas B, yang sudah bersedia mengisi dan menjadi bagian dalam cerita perkuliahan penulis.
10. Teman-teman asisten laboratorium algoritma dan pemograman yang memberikan banyak pengalaman berharga dalam sisi lain perkuliahan yakni Vivi, Andi, Irul, dan Raharjo.
11. Teman-teman BEM Fakultas Ilmu Komputer 2017-2018
12. Teman-teman UK Pers Mahasiswa, khususnya pengurus 2016, Hani, Arum, Kholis, Ilham, Radit, Inggrid, Silvi, Bachtiar, Fitri, dan Miya yang sudah banyak memberikan cerita semasa kuliah.
13. Teman-teman KKN 2017, khususnya group SSI, Stefany, Sonia, Rica, Sam, dan Fau yang juga turut memberi dorongan kepada penulis untuk menyelesaikan skripsi ini.
14. Tika, Linda, & Yasina, teman-teman baik saya yang selalu hadir dalam segala kondisi termasuk menyemangati penulis agar dapat menyelesaikan skripsi ini.
15. Bebe dan Mbak Dai, tim kamseupay, yang selalu memberi energi-energi positif dan mendukung penulis dalam segala keadaan.
16. Semua pihak yang tidak dapat penulis ucapkan yang juga turut mendoakan dan menyemangati penulis dalam menyelesaikan skripsi ini.

DAFTAR ISI

DAFTAR GAMBAR, GRAFIK, dan DIAGRAM

DAFTAR TABEL

DAFTAR SIMBOL