

**IMPLEMENTASI DASHBOARD WEB SEARCH ENGINE
UNTUK PEMROSESAN QUERY MENGGUNAKAN
ALGORITMA ENHANCED CONFIX STRIPPING DAN
ALGORITMA TF-IDF**

SKRIPSI



IKA SULISTIORINI

**PROGRAM STUDI S1-MATEMATIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**IMPLEMENTASI DASHBOARD WEB SEARCH ENGINE
UNTUK PEMROSESAN QUERY MENGGUNAKAN
ALGORITMA ENHANCED CONFIX STRIPPING DAN
ALGORITMA TF-IDF**

SKRIPSI

Sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana Sains Bidang Matematika
pada Fakultas Sains dan Teknologi
Universitas Airlangga

Ika Sulistiorini
NIM 081711233032

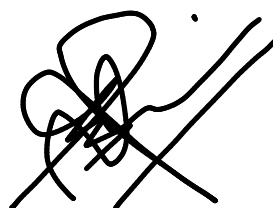
Disetujui oleh:

Pembimbing I,



Auli Damayanti, S.Si., M. Si.
NIP. 197511072003122004

Pembimbing II,



Drs. Edi Winarko, M.Cs.
NIP. 196705141993031001



LEMBAR PENGESAHAN NASKAH SKRIPSI

Judul : Implementasi Dashboard Web Search Engine Untuk
Pemrosesan Query Menggunakan Algoritma Enhanced
Confix Stripping dan Algoritma TF-IDF

Penyusun : Ika Sulistiorini

NIM : 081711233032

Tanggal Seminar : 12 Januari 2022

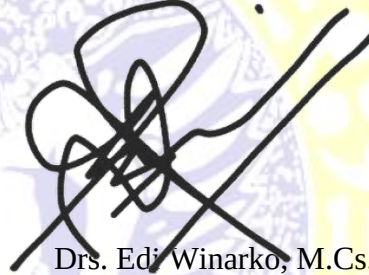
Disetujui Oleh:

Pembimbing I,



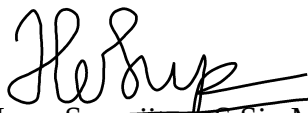
Auli Damayanti, S.Si., M. Si.
NIP. 197511072003122004

Pembimbing II,



Drs. Edi Winarko, M.Cs.
NIP. 196705141993031001

Mengetahui,
Ketua Departemen Matematika
Fakultas Sains dan Teknologi
Universitas Airlangga



Dr. Herry Suprajitno, S.Si., M.Si.
NIP. 196804041994031020

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan, namun tersedia di perpustakaan dalam lingkungan Universitas Airlangga, diperkenankan untuk dipakai sebagai referensi kepustakaan, tetapi pengutipan harus seizin penyusun dan harus menyebutkan sumbernya sesuai kebiasaan ilmiah. Dokumen skripsi ini merupakan hak milik Universitas Airlangga



SURAT PERNYATAAN TENTANG ORISINALITAS

Yang bertanda tangan dibawah ini, saya:

Nama : Ika Sulistiorini
NIM : 081711233032
Program Studi : S-1 Matematika
Fakultas : Sains dan Teknologi
Jenjang : Sarjana (S1)

Menyatakan bahwa saya tidak melakukan tindakan plagiat dalam penulisan skripsi saya yang berjudul:

“Implementasi Dashboard Web Search Engine untuk Pemrosesan Query Menggunakan Algoritma Enhanced Confix Stripping dan Algoritma Tf-Idf”

Apabila suatu saat nanti terbukti melakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan.

Demikian surat ini saya buat dengan sebenar – benarnya.

Surabaya, Januari 2022



METERAI
TEMPEL
825AJX520858129

Ika Sulistiorini

NIM. 081711233032

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah memberikan rahmat dan karunia-Nya sehingga kami dapat menyelesaikan skripsi yang berjudul “Implementasi Dashboard Web Search Engine untuk Pemrosesan Query Menggunakan Algoritma Enhanced Confix Stripping dan Algoritma Tf-Idf” dengan baik guna menyelesaikan mata kuliah skripsi serta mampu mengimplementasikan ilmu-ilmu yang telah diajarkan di Program Studi Matematika.

Skripsi ini dapat disusun dengan baik berkat bantuan dari pihak-pihak yang telah memberikan bimbingan dan dukungan sebagai bahan masukan untuk kami. Untuk itu pada kesempatan ini kami mengucapkan terima kasih kepada :

1. Prof Dr. Moh. Nasih SE.,MT.,Ak. selaku Rektor Universitas Airlangga.
2. Prof. Dr. Moh Yasin, M.Si selaku Dekan Fakultas Sains dan Teknologi Universitas Airlangga.
3. Dr. Herry Suprajitno, S.Si, M.Si. selaku Ketua Departemen Matematika sekaligus Koordinator Program Studi S1-Matematika Fakultas Sains dan Teknologi Universitas Airlangga.
4. Dr. Nenik Estuningsih, S.Si., M.Si. selaku dosen wali yang senantiasa memberikan saran dan arahan kepada penulis.
5. Auli Damayanti, S.Si., M.Si. sebagai Dosen Pembimbing I dan Drs. Edi Winarko, M.Cs. selaku pembimbing II yang telah sabar membimbing dan memberikan wawasan kepada penulis untuk menyelesaikan skripsi.
6. Kedua orang tua penulis, Bapak Suprpto dan Ibu Yatemi yang senantiasa tiada henti memberikan dukungan, motivasi, dan do'a untuk kesuksesan penulis.
7. Seluruh dosen dan staf di Fakultas Sains dan Teknologi Universitas Airlangga, khususnya di Departemen Matematika yang telah mengajarkan ilmu yang bermanfaat kepada penulis selama masa perkuliahan.

8. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang terlibat dalam membantu penyusunan skripsi baik secara langsung maupun tidak langsung.

Akhirnya, penulis mengucapkan terima kasih atas segala dukungan dan bantuan sehingga skripsi ini dapat tersusun dengan baik. Semoga skripsi ini dapat bermanfaat bagi penulis dan pembaca.

Surabaya, 3 Januari 2022

Penulis



Ika Sulistiorini, 2022, **Implementasi Dashboard Web Search Engine Untuk Pemrosesan Query Menggunakan Algoritma Enhanced Confix Stripping dan Algoritma TF-IDF**. Skripsi ini dibawah bimbingan Auli Damayanti, S.Si., M. Si. dan Drs. Edi Winarko, M.Cs. Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya.

ABSTRAK

Dashboard merupakan salah satu alat atau media yang dapat menjembatani pengguna dalam menemukan informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi dengan memasukkan suatu query sehingga memudahkan pengguna untuk mendapatkan informasi. Pada penulisan skripsi ini, dirancang sebuah program berupa dashboard web search engine untuk melakukan pencarian artikel dengan menerapkan algoritma *Stemming Enhanced Confix Stripping* sebagai algoritma stemming untuk mendapatkan bentuk kata dasar dari suatu kata dan algoritma TF-IDF untuk melakukan pembobotan. Langkah – langkah perancangan dashboard web search engine adalah melakukan text mining, pembobotan dokumen, dan pembangunan dashboard. Tahapan pada proses text mining adalah proses *tokenizing*, *filtering*, dan *stemming*. Proses *stemming* untuk setiap bahasa berbeda-beda, *stemming* kata berbahasa Indonesia dapat dilakukan dengan mengaplikasikan beberapa algoritma, Salah satu algoritmanya adalah *Enhanced Confix Stripping* (ECS), Algoritma ECS melakukan pencocokan kata hasil filtering dengan kamus kata dasar bahasa Indonesia kemudian melakukan penghilangan akhiran dan awalan sesuai dengan aturan pemenggalan awalan dan akhiran pada algoritma ECS sehingga diperoleh kata hasil stemming yang terbaik. Pada proses pembobotan dokumen digunakan algoritma TF-IDF yaitu dengan melakukan pembobotan query untuk setiap dokumen. Hasil pembobotan pada setiap dokumen diurutkan dan ditampilkan sebagai keluaran program. Proses pembangunan dashboard web search engine menggunakan bahasa pemrograman R yang diimplementasikan menggunakan shiny dashboard. Data yang digunakan berupa artikel ilmiah yang berjumlah 114 data, dengan rincian judul, pengarang, tahun, abstrak, dan link. Pada implementasinya, dashboard mampu menampilkan dokumen- dokumen yang banyak mengandung query yang diinputkan dengan bobot TF-IDF setiap dokumen lebih dari 3.

Kata Kunci: *Query, Text Mining, Tokenizing, Filtering, Stemming, Enhanced Confix Stripping, TF-IDF, Dashboard*

Ika Sulistiorini, 2022, **Implementation of Dashboard Web Search Engine for Query Processing Using Enhanced Confix Stripping Algorithm and TF-IDF Algorithm**. This thesis was supervised by Auli Damayanti, S.Si., M. Si. and Drs. Edi Winarko, M.Cs. Department of Mathematics, Faculty of Science and Technology, Airlangga University, Surabaya.

ABSTRACT

Dashboard is a tool or media that can bridge users in finding information relevant to user needs from a collection of information by entering a query to make it easier for users to get information. In this thesis, designed a program in the form of a web dashboard search engine to search articles by applying *Stemming Enhanced Confix Stripping* algorithm as stemming algorithms to get the basis of a word and TF-IDF algorithm to perform weighting. The steps for designing a web search engine dashboard are text mining, document weighting, and dashboard construction. The stages in the text mining process are *tokenizing*, *filtering*, and *stemming*. The *stemming* process for each language is different, *stemming* Indonesian word can be done by applying several algorithms, one of the algorithms is *Enhanced Confix Stripping* (ECS), the ECS algorithm matches the filtered words with the basic Indonesian dictionary and then removes suffixes and prefixes according to the rules of beheading the prefix and suffix on the ECS algorithm so that the best stemming results are obtained. In the document weighting process, the TF-IDF algorithm is used, namely by weighting the query for each document. The results of the weighting of each document are sorted and displayed as program output. The process of building a web search engine dashboard uses the R programming language which is implemented using a shiny dashboard. The data used in the form of scientific articles totaling 114 data, with details of the title, author, year, abstract, and link. In its implementation, the dashboard is able to display documents that contain a lot of inputted queries with a TF-IDF weight of more than 3.

Keywords: *Query, Text Mining, Tokenizing, Filtering, Stemming, Enhanced Confix Stripping, TF-IDF, Dashboard*

DAFTAR ISI

JUDUL	i
LEMBAR PENGESAHAN NASKAH SKRIPSI.....	iv
PEDOMAN PENGGUNAAN SKRIPSI	v
SURAT PERNYATAAN TENTANG ORISINALITAS	vi
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT.....	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Web Search Engine	6
2.2 Dashboard.....	7
2.3 Shiny Dashboard	8
2.4 Text Mining.....	10
2.5 Stemming.....	10
2.6 Algoritma Enhanced Confix Stripping.....	11
2.7 Algoritma TF-IDF	16
2.8 Cosine Similarity	17

2.9	Peringkasan Dokumen Menggunakan TF-IDF	18
BAB III METODOLOGI PENELITIAN.....		20
BAB IV PEMBAHASAN.....		25
4.1	Pengambilan Data.....	25
4.2	Prosedur Pengolahan Dokumen	27
4.2.1	Tahap pengambilan data dari database	27
4.2.2	Tahap tokenizing.....	28
4.2.3	Tahap Filtering.....	30
4.2.4	Tahap Stemming	31
4.2.5	Perhitungan TF-IDF.....	32
4.2.6	Peringkasan Abstrak Dokumen dengan TF-IDF	33
4.3	Perhitungan Manual	35
4.3.1	Inisialisasi Kata Kunci	35
4.3.2	Proses Tokenizing.....	35
4.3.3	Proses Filtering	37
4.3.4	Proses Stemming.....	38
4.3.5	Proses Perhitungan TF-IDF	40
4.3.6	Peringkasan dokumen menggunakan TF-IDF	45
4.4	Implementasi program.....	52
BAB V KESIMPULAN DAN SARAN.....		62
5.1	Kesimpulan.....	62
5.2	Saran.....	63
DAFTAR PUSTAKA		64
LAMPIRAN.....		67

DAFTAR GAMBAR

Gambar 2.1 Tampilan Header Shiny Dashboard.....	8
Gambar 2.2 Tampilan Sidebar Shiny Dashboard	9
Gambar 2.3 Tampilan Body Shiny Dashboard.....	9
Gambar 2.4 Ilustrasi cosine similarity	17
Gambar 4. 1 Struktur db_skripsi	25
Gambar 4. 2 Prosedur pengambilan dan penyimpanan data	27
Gambar 4. 3 Prosedur cleaning data	28
Gambar 4. 4 Tokenizing keyword	29
Gambar 4. 5 Tokenizing data	29
Gambar 4. 6 Filtering data.....	30
Gambar 4. 7 Filtering keyword	30
Gambar 4. 8 Prosedur stemming dengan algoritma ECS	31
Gambar 4. 9 Prosedur pengurutan hasil perhitungan TF-IDF	32
Gambar 4. 10 Prosedur Tokenizing Kalimat.....	33
Gambar 4. 11 Prosedur TF-IDF Peringkasan Dokumen	34
Gambar 4. 12 Tampilan Awal Dashboard	53
Gambar 4. 13 Tampilan Menu Dashboard Pencarian.....	53
Gambar 4. 14 Tampilan Menu Hasil Pencarian	54
Gambar 4. 15 Tampilan Menu Analisis 1.....	54
Gambar 4. 16 Tampilan Menu Analisis 2.....	55
Gambar 4. 17 Tampilan Menu Analisis 3.....	55
Gambar 4. 18 Tampilan Menu Peringkasan Dokumen	56
Gambar 4. 19 Tampilan Menu Login Admin.....	56
Gambar 4. 20 Tampilan Dashboard Admin.....	57
Gambar 4. 21 Tampilan Modifikasi Dokumen.....	57
Gambar 4. 22 Tampilan Detail Dokumen	58
Gambar 4. 23 Tampilan Menu Tambah Dokumen.....	58
Gambar 4. 24 Tampilan Daftar Admin.....	59
Gambar 4. 25 Tampilan Detail Admin	59

Gambar 4. 26 Tampilan Modifikasi Admin	60
Gambar 4. 27 Tampilan Tambah Admin.....	60
Gambar 4. 28 Tampilan Menu About.....	61



DAFTAR TABEL

Tabel 2. 1 Aturan pemenggalan awalan pada algoritma ECS	13
Tabel 4. 1 Hasil tokenizing kata kunci	35
Tabel 4. 2 Hasil filtering kata kunci	37
Tabel 4. 3 Hasil stemming kata kunci	38
Tabel 4. 4 Hasil perhitungan TF kata kunci	40
Tabel 4. 5 Hasil perhitungan IDF kata kunci	41
Tabel 4. 6 Hasil perhitungan TF-IDF	42
Tabel 4. 7 Perhitungan cosine similarity	43
Tabel 4. 8 Pemecahan abstrak menjadi kalimat	46
Tabel 4. 9 Perhitungan Term Frequency untuk setiap kata unik	47
Tabel 4. 10 Perhitungan TF-IDF untuk setiap kata unik	48

DAFTAR LAMPIRAN

Lampiran 1. Flowchart System.....	67
Lampiran 2. Flowchart stemming	68
Lampiran 3. Flowchart TF-IDF	69
Lampiran 4 Prosedur pengolahan dokumen.....	70
Lampiran 5. Prosedur perhitungan TF-IDF	72
Lampiran 6. Prosedur perhitungan cosine similarity	73
Lampiran 7. Abstrak Dokumen 1.....	74
Lampiran 8. Abstrak Dokumen 2.....	75
Lampiran 9 . Abstrak Dokumen 3.....	75
Lampiran 10. Abstrak Dokumen 4.....	76
Lampiran 11. Abstrak Dokumen 5.....	76
Lampiran 12. Abstrak Dokumen 6.....	77
Lampiran 13. Abstrak Dokumen 7.....	78
Lampiran 14. Abstrak Dokumen 8.....	79
Lampiran 15. Tokenizing Dokumen 1.....	80
Lampiran 16. Tokenizing Dokumen 2.....	81
Lampiran 17. Tokenizing Dokumen 3.....	82
Lampiran 18. Tokenizing Dokumen 4.....	83
Lampiran 19. Tokenizing Dokumen 5.....	84
Lampiran 20. Tokenizing Dokumen 6.....	86
Lampiran 21. Tokenizing Dokumen 7.....	87
Lampiran 22. Tokenizing Dokumen 8.....	88
Lampiran 23. Filtering Dokumen 1.....	90
Lampiran 24. Filtering Dokumen 2.....	91
Lampiran 25. Filtering Dokumen 3.....	92
Lampiran 26. Filtering Dokumen 4.....	93
Lampiran 27. Filtering Dokumen 5.....	94
Lampiran 28. Filtering Dokumen 6.....	95
Lampiran 29. Filtering Dokumen 7.....	96
Lampiran 30. Filtering Dokumen 8.....	97
Lampiran 31. Stemming Dokumen 1	98
Lampiran 32. Stemming Dokumen 2	99

Lampiran 33. Stemming Dokumen 3	100
Lampiran 34. Stemming Dokumen 4	101
Lampiran 35. Stemming Dokumen 5	102
Lampiran 36. Stemming Dokumen 6	103
Lampiran 37. Stemming Dokumen 7	104
Lampiran 38. Stemming Dokumen 8	105
Lampiran 39. Daftar Stopword Bahasa Indonesia	106

