

**IMPLEMENTASI MODUL INDEXING PADA SEARCH ENGINE  
TELUSURI DENGAN INTEGRASI INVERTED INDEX DAN  
GENERALIZED SUFFIX TREE UNTUK MEREDUKSI WAKTU  
PENCARIAN**

**Skripsi**

**Disusun untuk memenuhi salah satu syarat  
memperoleh gelar Sarjana Komputer**



*Mencerdaskan dan  
Memartabatkan Bangsa*

**Oleh:  
Mochammad Hanif Ramadhan  
1313619025**

**PROGRAM STUDI ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI JAKARTA**

**2023**

## ABSTRAK

**MOCHAMMAD HANIF RAMADHAN.** Implementasi Modul Indexing Pada Search Engine Telusuri Dengan Integrasi Inverted Index Dan Generalized Suffix Tree Untuk Mereduksi Waktu Pencarian. Skripsi. Program Studi Ilmu Komputer. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. Agustus 2023. Di bawah bimbingan Muhammad Eka Suryana, M.Kom dan Med Irzal, M.Kom.

Mesin pencari atau *search engine* adalah program komputer yang digunakan untuk melakukan pencarian situs web. Pencarian dapat dilakukan dengan mengumpulkan informasi tentang halaman web terlebih dahulu. Karena ukuran data yang besar sebagai sumber informasi utama bagi mesin pencari, penggunaan *index* bisa dimanfaatkan untuk mereduksi waktu pencarian. Penelitian ini merupakan bagian dari rangkaian penelitian mesin pencari *Telusuri*, dan bertujuan untuk membuat implementasi *index* berdasarkan struktur *inverted index*, yaitu struktur *index* yang digunakan oleh mesin pencari *Google*, dan melakukan integrasi dengan struktur *Generalized Suffix Tree*. Tujuan utamanya adalah untuk mereduksi waktu pencarian suatu *keyword* dan memberikan peringkat terhadap dokumen berdasarkan kesesuaian antara informasi dalam dokumen dengan teks yang dimasukkan oleh pengguna. Hasil akhir dari implementasi modul menunjukkan reduksi waktu yang signifikan, berkisar antara 89 - 98%.

**Kata kunci:** mesin pencari, indeks, basis data, teori informasi

## ABSTRACT






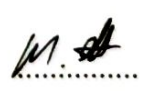

**MOCHAMMAD HANIF RAMADHAN.** Implementasi Modul Indexing Pada Search Engine Telusuri Dengan Integrasi Inverted Index Dan Generalized Suffix Tree Untuk Mereduksi Waktu Pencarian. Mini Thesis. Computer Science. Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta. August 2023. Under guidance from Muhammad Eka Suryana, M.Kom and Med Irzal, M.Kom.

Search engine is a computer program that is used to search for a webpage. Before searching can be done, it is required to gather informatoin about the webpage first. Due to the large collection of data needed for a search engine to be usable, indexes can be used to reduce time to retrieve information. This research is a part of a longer research for *Telusuri* search engine, and aims to implement an index structure based on inverted index form, which are used by *Google*, and integrates it with the *Generalized Suffix Tree*. The goals is to reduce the time taken to get informations from a given keyword and ranks the result based on the relevancy between the webpage and the input keyword. The index implementations are shown to be able to reduce the time taken significantly, ranging from 89 to 98%.

**Kata kunci:** *search engine, indexing, database, information retrieval, tree*

**LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI**  
**IMPLEMENTASI MODUL INDEXING PADA SEARCH ENGINE**  
**TELUSURI DENGAN INTEGRASI INVERTED INDEX DAN**  
**GENERALIZED SUFFIX TREE UNTUK MEREDUKSI WAKTU**  
**PENCARIAN**

**Nama : Mochammad Hanif Ramadhan**  
**No. Registrasi : 1313619025**

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: <u>Prof. Dr. Muktiningsih N. M.Si.</u> NIP. 196405111989032001		1-09-2023
Wakil Penanggung Jawab			
Wakil Dekan I	: <u>Dr. Esmar Budi, S.Si., MT.</u> NIP. 197207281999031002		1-09-2023
Ketua	: <u>Drs. Mulyono, M.Kom.</u> NIP. 19660517 199403 1 003		23-08-2023
Sekretaris	: <u>Ari Hendarno, S.Pd., M.Kom.</u> NIP. 19881102 202203 1 002		23-08-2023
Penguji	: <u>Dr. Ria Arafiyah, M.Si.</u> NIP. 19751121 200501 2 004		23-08-2023
Pembimbing I	: <u>Muhammad Eka Suryana, M.Kom.</u> NIP. 19851223 201212 1 002		22-08-2023
Pembimbing II	: <u>Med Irzal, M.Kom</u> NIP. 19770615 200312 1 001		23-09-2023

Dinyatakan lulus UJIAN SKRIPSI tanggal: 21 Agustus 2023



## LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul **"Implementasi Modul Indexing Pada Search Engine Telusuri Dengan Integrasi Inverted Index Dan Generalized Suffix Tree Untuk Mereduksi Waktu Pencarian"** yang disusun sebagai syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Ilmu Komputer Universitas Negeri Jakarta adalah karya ilmiah saya dengan arahan dari dosen pembimbing.

Sumber informasi yang diperoleh dari peneliti lain yang telah dipublikasikan yang disebutkan dalam teks Skripsi ini, telah dicantumkan dalam Daftar Pustaka sesuai dengan norma, kaidan dan etika penulisan ilmiah.

Jika dikemudian hari ditemukan sebagian besar skripsi ini bukan hasil karya saya sendiri dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sanding dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku.

Jakarta, 21 Agustus 2023



Mochammad Hanif Kamadhan



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN  
UNIVERSITAS NEGERI JAKARTA  
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220  
Telepon/Faksimili: 021-4894221  
Laman: [lib.unj.ac.id](http://lib.unj.ac.id)

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : MOCHAMMAD HANIF RAMADHAN  
NIM : 1313619025  
Fakultas/Prodi : FMIPA / ILMU KOMPUTER  
Alamat email : [eternal70@gmail.com](mailto:eternal70@gmail.com)

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi  Tesis  Disertasi  Lain-lain (.....)

yang berjudul :

IMPLEMENTASI MODUL INDEXING PADA SEARCH ENGINE TELUSURI DENGAN  
INTEGRASI INVERTED INDEX DAN GENERALIZED SUFFIX TREE UNTUK MEREDUKSI  
WAKTU PENCARIAN

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 2 September 2023

Penulis

( MOCHAMMAD HANIF R )  
nama dan tanda tangan

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, karena dengan rahmat dan karunia-Nya, penulis dapat menyelesaikan proposal skripsi yang berjudul *Implementasi Modul Indexing Pada Search Engine Telusuri Dengan Integrasi Inverted Index Dan Generalized Suffix Tree Untuk Mereduksi Waktu Pencarian*.

Keberhasilan dalam penyusunan proposal skripsi ini tidak lepas dari bantuan berbagai pihak yang mana dengan tulus dan ikhlas memberikan masukan guna sempurnanya proposal skripsi ini. Oleh karena itu dalam kesempatan ini, dengan kerendahan hati penulis mengucapkan banyak terima kasih kepada:

1. Yth. Para petinggi di lingkungan FMIPA Universitas Negeri Jakarta.
2. Yth. Ibu Dr. Ria Arafiyah, M.Si selaku Koordinator Program Studi Ilmu Komputer.
3. Yth. Bapak Muhammad Eka Suryana, M.Kom selaku Dosen Pembimbing I yang telah membimbing, mengarahkan, serta memberikan saran dan koreksi terhadap proposal skripsi ini.
4. Yth. Bapak Med Irzal, M.Kom selaku Dosen Pembimbing II yang telah membimbing, mengarahkan, serta memberikan saran dan koreksi terhadap proposal skripsi ini.
5. Kedua orang tua dan kakak penulis yang telah mendukung dan memberikan semangat serta doa untuk penulis.
6. Wulan sebagai seseorang yang mau mendengarkan keluh kesah dan mendukung penulis.
7. Teman-teman Warung Tegang yang walau dengan segala keanehannya selalu memberikan semangat dan motivasi bagi penulis
8. Teman-teman SMAN 68 yang seringkali menemani penulis bertukar pikiran ketika beristirahat sejenak selama penulisan
9. Teman-teman Program Studi Ilmu Komputer 2019 yang telah memberikan dukungan dan memiliki andil dalam penulisan proposal skripsi ini.

10. Kedua kucing milik Pak Eka yang seringkali menghibur penulis selama melaksanakan bimbingan di Jasinga

Penulis menyadari bahwa penyusunan proposal skripsi ini masih jauh dari sempurna karena keterbatasan ilmu dan pengalaman yang dimiliki. Oleh karenanya, kritik dan saran yang bersifat membangun akan penulis terima dengan senang hati. Akhir kata, penulis berharap tugas akhir ini bisa bermanfaat bagi semua pihak khususnya penulis sendiri. Semoga Allah SWT senantiasa membalas kebaikan semua pihak yang telah membantu penulis dalam menyelesaikan proposal skripsi ini.

Jakarta, 21 Agustus 2023

Mochammad Hanif Ramadhan





# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>iv</b>
<b>ABSTRAK</b>	<b>v</b>
<b>ABSTRACT</b>	<b>vi</b>
<b>DAFTAR ISI</b>	<b>vii</b>
<b>DAFTAR GAMBAR</b>	<b>ix</b>
<b>I PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang Masalah . . . . .	1
1.2 Rumusan Masalah . . . . .	4
1.3 Batasan Masalah . . . . .	4
1.4 Tujuan Penelitian . . . . .	4
1.5 Manfaat Penelitian . . . . .	4
<b>DAFTAR TABEL</b>	<b>1</b>
<b>II KAJIAN PUSTAKA</b>	<b>6</b>
2.1 Proses <i>Indexing</i> . . . . .	6
2.2 <i>Repository</i> . . . . .	7
2.3 Daftar URL . . . . .	7
2.4 <i>Document Index</i> . . . . .	7
2.5 Daftar Kosakata . . . . .	8
2.6 <i>docID</i> dan <i>wordID</i> . . . . .	8
2.7 <i>Inverted Index</i> . . . . .	8
2.7.1 Struktur Index . . . . .	9
2.7.2 Skema Penyimpanan Index . . . . .	11
2.7.3 Pemeringkatan Query . . . . .	11
2.7.4 <i>Similarity Scoring</i> . . . . .	12
2.8 Proses Pembuatan Index . . . . .	12
2.8.1 Pengolahan Query . . . . .	14
2.9 <i>Jaccard Distance</i> . . . . .	15
2.10 Komunikasi antar proses . . . . .	16
2.11 Modul <i>typing</i> pada Python . . . . .	17
2.12 Perbandingan Metode <i>Indexing</i> . . . . .	17
2.12.1 <i>Signature File Index</i> . . . . .	17
2.12.2 <i>Bitmap Index</i> . . . . .	18
2.12.3 <i>WordNet</i> . . . . .	18

<b>III METODOLOGI PENELITIAN</b>	<b>19</b>
3.1 Tahapan Penelitian . . . . .	19
3.2 Konstruksi Index . . . . .	20
3.3 Daftar kata umum . . . . .	22
3.4 Metode Pemeringkatan . . . . .	23
3.4.1 <i>Word distance</i> ( $\gamma$ ) . . . . .	23
3.4.2 <i>Word similarity</i> ( $\beta$ ) . . . . .	24
3.5 Modifikasi Arsitektur <i>Telusuri</i> . . . . .	25
3.6 Skema Pencarian . . . . .	26
3.7 Alat dan Bahan Penelitian . . . . .	30
3.8 Tahapan Pengembangan . . . . .	30
3.8.1 Meningkatkan kemampuan modul <i>indexing</i> saat ini . . . . .	30
3.8.2 Rancangan Eksperimen . . . . .	31
<b>IV HASIL DAN PEMBAHASAN</b>	<b>33</b>
4.1 Implementasi . . . . .	33
4.1.1 Pengolahan <i>Dataset</i> . . . . .	34
4.1.2 Pembentukan <i>Inverted Index</i> . . . . .	35
4.1.3 Pengolahan <i>User Query</i> . . . . .	38
4.1.4 Proses Penggabungan <i>Hitlist</i> dan Pemeringkatan Hasil . . . . .	41
4.1.5 Integrasi dengan <i>Generalized Suffix Tree (GST)</i> . . . . .	45
4.1.6 Modifikasi kode <i>TF-IDF</i> . . . . .	49
4.1.7 Struktur Direktori Kode . . . . .	54
4.2 Pengujian . . . . .	54
4.2.1 Statistik Pengujian . . . . .	54
4.2.2 Pengujian Relevansi . . . . .	56
4.2.3 Pengujian Performa . . . . .	64
4.3 Analisis Hasil . . . . .	64
<b>V KESIMPULAN DAN SARAN</b>	<b>66</b>
5.1 Kesimpulan . . . . .	66
5.2 Saran . . . . .	66
<b>DAFTAR PUSTAKA</b>	<b>68</b>
<b>LAMPIRAN</b>	<b>69</b>
<b>A Dokumentasi Pengujian Metode <i>Inverted Index</i></b>	<b>69</b>
<b>B Dokumentasi Pengujian Metode <i>Inverted Index</i> dan Integrasi <i>GST</i></b>	<b>71</b>