



**MENINGKATKAN EFEKTIFITAS DALAM *TEXT*
PREPROCESSING DENGAN IMPLEMENTASI
ELECTRONIC DATA PROCESSING PADA TEKS
BERITA**

SKRIPSI

Rafdi Reyhan Zhafari

1810511036

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAKARTA**

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 INFORMATIKA

2023



**MENINGKATKAN EFEKTIFITAS DALAM *TEXT*
PREPROCESSING DENGAN IMPLEMENTASI
ELECTRONIC DATA PROCESSING PADA TEKS
BERITA**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Komputer**

Rafdi Reyhan Zhafari

1810511036

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN
JAKARTA**

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI S1 INFORMATIKA

2023

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun dirujuk saya nyatakan dengan benar.

Nama : Rafdi Reyhan Zhafari

NIM : 1810511036

Tanggal : 13 Juli 2023

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta, 13 Juli 2023

Yang Menyatakan,



Rafdi Reyhan Zhafari

PERYATAAN PERSETUJUAN PUBLIKASI

Sebagai *civitas* akademik universitas Pembangunan Nasional Veteran Jakarta, saya bertanda tangan di bawah ini:

Nama : Rafdi Reyhan Zhafari

NIM : 1810511036

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-Exclusive Royalty Free Right*) atas karya ilmiah yang berjudul:

Meningkatkan Efektifitas Dalam Text Preprocessing Dengan Implementasi Electronic Data Processing Pada Teks Berita

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih formatkan, mengelola dalam bentuk pangkalan data, merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 13 Juli 2023

Yang Menyatakan



Rafdi Reyhan Zhafari

LEMBAR PENGESAHAN

Dengan ini menyatakan bahwa Skripsi berikut:

Nama : Rafdi Reyhan Zhafari

NIM : 1810511036

Program Studi : S1 Informatika

Judul Skripsi : Meningkatkan Efektifitas Dalam Text Preprocessing Dengan Implementasi Electronic Data Processing Pada Teks Berita

Telah berhasil dipertahankan dihadapan tim penguji dan diterima sebagai persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.



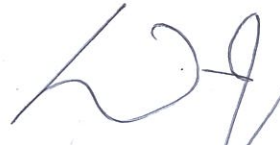
Dr. Ermatita, M.Kom.
Penguji Satu



Bayu Hananto, S.Kom, M.Kom.
Penguji Dua



Iin Ernawati, S.Kom, M.Si.
Pembimbing



Dr. Widya Cholil, S.Kom., M.I.T.
Ketua Program Studi



Dr. Ermatita, M.Kom.
Dekan

Ditetapkan di : Jakarta

Tanggal Persetujuan : 11 Juli 2023



MENINGKATKAN EFEKTIFITAS DALAM TEXT PREPROCESSING DENGAN IMPLEMENTASI ELECTRONIC DATA PROCESSING PADA TEKS BERITA

Rafdi Reyhan Zhafari

ABSTRAK

Tersebarnya informasi yang cepat dan masif tentunya membutuhkan pemrosesan yang cepat dan masif pula. Seiring berjalannya waktu teknologi menjadi suatu hal yang rumit untuk disaring atau disarikan dari data awalnya, hal ini lah yang membuat electronic data processing menjadi sebuah hal secara tidak disengaja maupun sengaja harus diadaptasi oleh manusia agar tidak tertinggal dari mendapatkan informasi yang dibutuhkan. Pada penelitian ini dilakukan electronic data processing pada text processing dengan menggunakan teks berita. Pada penelitian dilakukan perbandingan dengan menggunakan perangkat lain, di mana dengan perangkat pada penelitian dihasilkan waktu 18,917 detik pada tahap preprocessing sedangkan pada perangkat colab research didapatkan hasil 17,431 detik.

Kata kunci: *electronic data processing, teks, preprocessing.*

IMPROVING EFFECTIVENESS IN TEXT PREPROCESSING WITH ELECTRONIC DATA PROCESSING IMPLEMENTATION IN NEWS TEXT

Rafdi Reyhan Zhafari

ABSTRACT

The rapid and massive dissemination of information certainly requires fast and massive processing as well. Over time, technology has become a complicated matter to filter or abstract from the initial data, this is what makes electronic data processing a thing that humans accidentally or intentionally have to adapt so they are not left behind from getting the information they need. In this research, electronic data processing was carried out on text processing using news text. In the study, a comparison was made using other devices, where with the device in the study, the result was 18.917 seconds at the preprocessing stage, while the results for the colab research device were 17.431 seconds.

Keyword : *electronic data processing, teks, preprocessing.*

KATA PENGANTAR

Pengantar ini penulis awali dengan mengucapkan syukur kepada Allah Tuhan semesta alam atas segala nikmat yang diberikan sehingga penulis dapat menyelesaikan Skripsi dengan judul “Meningkatkan Efektifitas Dalam Text Preprocessing Dengan Implementasi Electronic Data Processing Pada Teks Berita”.

Shalawat dan salam semoga tersampaikan kepada Nabi Muhammad ﷺ, keluarga dan para Sahabat beliau, serta orang-orang yang meneladani mereka sampai hari akhir nanti.

Rasa terima kasih yang banyak penulis sampaikan kepada keluarga yang tercinta, teman-teman yang penulis banggakan, dan saudara-saudara yang baik secara langsung maupun tidak langsung telah menjadi bagian dari tersusunnya tulisan ini. Sejak hari pertama penulisan ini tidak ada yang berhenti mendukung baik secara moral maupun materi.

Tidak lupa penulis sampaikan jazakumullahu khairan kepada seluruh dosen yang baik secara langsung maupun tidak telah mengajarkan kebaikan yang luar biasa kepada penulis. Tidak ada ilmu yang tiap dosen sampaikan kecuali ilmu itu berbuah kebaikan untuk penulis.

Setiap tulisan ini tidak tertulis dengan sendirinya melainkan tulisan ini tertulis dengan rasa sadar dan tanggung jawab sebagai bagian dari syarat kelulusan dari perguruan tinggi negeri. Universitas Pembangunan Nasional Veteran Jakarta sudah menjadi tempat yang nyaman untuk terciptanya tulisan ini sehingga tulisan ini dapat diselesaikan sebagaimana mestinya.

Akhir kata penulis berharap semoga tulisan ini tidak berhenti hanya pada sebuah memori tetapi tetap berguna untuk dapat diteruskan dan dikembangkan untuk kebaikan bersama.

Jakarta, 13 Juli 2023

Rafdi Reyhan Zhafari

DAFTAR ISI

LEMBAR PERSETUJUAN.....	iii
ABSTRAK	iv
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	ix
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Ruang Lingkup	3
1.5 Luaran yang Diharapkan	3
1.6 Manfaat Penelitian.....	3
BAB 2 TINJAUAN PUSTAKA	4
2.1 Electronic Data Processing.....	4
2.2 Google Alerts.....	4
2.3 XML	6
2.4 Text Mining	7
2.5 Crawling	7
2.6 Preprocessing.....	8
2.6.1 Data Cleaning.....	9
2.6.2 Case Folding.....	9
2.6.3 Stopword Removal.....	9
2.6.4 Stemming	10

2.6.5	Tokenizing.....	11
2.7	Penelitian Terdahulu.....	11
BAB 3	METODE PENELITIAN.....	13
3.1	Kerangka Pikir.....	13
3.2	Identifikasi Masalah.....	14
3.3	Studi Literatur.....	14
3.4	Pengumpulan Data.....	14
3.5	<i>Preprocessing</i>	15
3.6	Dokumentasi.....	16
3.7	Perangkat Penelitian.....	16
3.7.1	Perangkat Keras (<i>hardware</i>).....	16
3.7.2	Perangkat Lunak (<i>software</i>).....	16
3.8	Jadwal Penelitian.....	17
BAB 4	HASIL DAN PEMBAHASAN.....	18
4.1	Pengumpulan Data.....	18
4.2	<i>Crawling</i> Data.....	20
4.3	<i>Preprocessing</i>	21
4.3.1	Case Folding.....	21
4.3.2	Symbol Removal.....	22
4.3.3	Stopword Removal.....	24
4.3.4	Tokenisasi.....	25
4.3.5	Stemming.....	27
4.4	Data Keluaran.....	28
BAB 5	PENUTUP.....	30
5.1	Kesimpulan.....	30

5.2	Saran.....	30
	DAFTAR PUSTAKA.....	32

DAFTAR GAMBAR

Gambar 2.1	Tampilan Awal Google Alerts.....	5
Gambar 2.2	Tampilan Query Terisi Google Alerts.....	5
Gambar 2.3	Tampilan Pilihan Pencarian Google Alerts.....	6
Gambar 3.1	Kerangka Pikir.....	13
Gambar 4.1	Google Alert Sharing Option.....	18
Gambar 4.2	Struktur File <i>XML</i>	19
Gambar 4.3	Gambar Tampilan Setelah Url Terisi	20

DAFTAR TABEL

Tabel 2.1	Penelitian Terdahulu	12
Tabel 3.1	Jadwal Penelitian	17
Tabel 4.1	Tabel Hasil Case Folding.....	22
Tabel 4.2	Tabel Hasil Symbol Removal	23
Tabel 4.3	Tabel Hasil Stopword Removal.....	24
Tabel 4.4	Tabel Hasil Tokenisasi	26
Tabel 4.5	Tabel Hasil Stemming	27
Tabel 4.6	Data Hasil Preprocessing.....	29