

**PENCARIAN DOKUMEN ARSIP DENGAN METODE
MULTICHARACTER AC-NFA BERBASIS GPU**

TUGAS AKHIR

Oleh:

THOMAS YAPUTRA (NIM. 141111201)

DARWIN SUCIPTA (NIM. 141110396)

RYAN DARIAN (NIM. 141110582)



UNIVERSITAS
MIKROSKIL

**PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
MIKROSKIL
MEDAN
2018**

**ARCHIVE DOCUMENT SEARCH WITH
MULTICHARACTER AC-NFA METHOD BASED ON GPU**

FINAL RESEARCH

By:

THOMAS YAPUTRA (NIM. 141111201)

DARWIN SUCIPTA (NIM. 141110396)

RYAN DARIAN (NIM. 141110582)



**UNIVERSITAS
MIKROSKIL**

**STUDY PROGRAM OF INFORMATICS ENGINEERING
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
MIKROSKIL**

MEDAN

2018

LEMBARAN PENGESAHAN
PENCARIAN DOKUMEN ARSIP DENGAN METODE
MULTICHARACTER AC-NFA BERBASIS GPU

TUGAS AKHIR

Diajukan untuk Melengkapi Persyaratan Guna
Mendapatkan Gelar Sarjana Strata Satu
Program Studi Teknik Informatika

Oleh :

THOMAS YAPUTRA (NIM. 141111201)

DARWIN SUCIPTA (NIM. 141110396)

RYAN DARIAN (NIM. 141110582)

Disetujui oleh :

Dosen Pembimbing I


(Arwin Halim, S.Kom., M.Kom)

Medan,.....21 Agustus.....2018

Diketahui dan Disahkan Oleh :

Ketua Program Studi

Teknik Informatika,



(Sunario Megawan, S.Kom., M.Kom)

ABSTRAK

Dokumen arsip memiliki peranan yang penting dalam kegiatan tata usaha suatu organisasi, dokumen tersebut biasanya tersimpan dalam bentuk *hardcopy* dan *softcopy*. Jumlah dokumen pada suatu organisasi pastinya tidak sedikit, oleh karena itu dibutuhkan sistem pencarian dokumen yang akan mencari dokumen sesuai *input user*. *Optical Character Recognition (OCR)* adalah proses pengenalan teks dalam gambar. Dengan menggunakan OCR maka proses pencarian dapat memperoleh dokumen yang sesuai dengan *input user*.

Metode yang dilakukan dalam tugas akhir ini adalah *Multicharacter AC-NFA* berbasis GPU yang akan mengekstrak teks dalam dokumen dengan library Tesseract-OCR dan melakukan pencarian dengan *multicharacter AC-NFA* secara paralel. Hasil pencarian tersebut akan disimpan dalam suatu variabel dan dokumen yang memiliki nilai dari variabel tersebut akan dimunculkan dalam hasil pencarian.

Pengujian dilakukan dengan melakukan perbandingan kecepatan pencarian antara CPU dengan GPU. Berdasarkan hasil pengujian terhadap sistem, diperoleh bahwa pencarian dengan CPU lebih cepat dibandingkan dengan pencarian GPU. Dengan catatan bahwa pada CPU kecepatan pencarian berbeda berdasarkan kata kunci yang digunakan sedangkan pada GPU kecepatan pencarian cenderung berada pada *range* kecepatan yang sama.

Kata kunci: Pencarian dokumen, *Optical Character Recognition (OCR)*, *Tesseract-OCR*, *Multicharacter AC-NFA*, *Aho-Corasick Algorithm*, *Graphics Processing Unit (GPU)*

KATA PENGANTAR

Puji dan syukur penulis ucapkan kepada Tuhan Yang Maha Esa, atas rahmat dan karunia-Nya penulis dapat menyelesaikan penulisan Tugas Akhir ini.

Adapun tujuan dari penulisan tugas akhir ini adalah sebagai salah satu syarat untuk menyelesaikan pendidikan guna mendapatkan gelar Sarjana Starata Satu Program Studi Teknik Informatika pada STMIK Mikroskil Medan.

Dalam penulisan tugas akhir ini, penulis telah berusaha semaksimal mungkin dan semua ini dapat tercapai berkat dukungan dan bantuan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini, penulis ingin menyampaikan terima kasih kepada yang terhormat :

1. Bapak Arwin Halim, S.Kom., M.Kom., selaku Pembimbing I.
2. Bapak Felix Pandi, S.Kom., M.Kom., selaku Pendamping Pembimbing.
3. Bapak Dr.Mimpin Ginting, M.S., selaku Ketua STMIK Mikroskil Medan.
4. Bapak Djoni, S.Kom., M.T.I., selaku Wakil Ketua I STMIK Mikroskil Medan.
5. Bapak Sunario Megawan S.Kom., M.Kom, selaku Ketua Program Studi Teknik Informatika STMIK Mikroskil Medan.
6. Seluruh dosen dan seluruh pegawai STMIK Mikroskil yang telah mengajar dan membantu selama perkuliahan.
7. Ayah, Ibu, serta teman-teman atas dukungan dan do'a yang telah diberikan.
8. Dan semua orang yang telah membantu penulis dalam menyelesaikan Tugas Akhir ini.

Penulis menyadari bahwa tugas akhir ini belum sempurna, akan tetapi penulis berharap tugas akhir ini dapat bermanfaat dan diterima bagi yang membutuhkan.

Medan, Agustus 2018

Thomas Yaputra

Darwin Sucipta

Ryan Darian

DAFTAR ISI

ABSTRAK	i
KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR	v
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	viii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan.....	2
1.4 Manfaat.....	2
1.5 Batasan Masalah	2
1.6 Metodologi Pengembangan Sistem	3
BAB II TINJAUAN PUSTAKA	5
2.1 Arsip	5
2.2 Information Retrieval	6
2.2.1 <i>Text-Based Image Retrieval</i> (TBIR).....	7
2.2.2 <i>Content-Based Image Retrieval</i> (CBIR).....	7
2.3 <i>String Matching</i>	8
2.3.1 Klasifikasi Algoritma <i>String Matching</i>	9
2.3.2 Teknik Algoritma <i>String Matching</i>	9
2.3.3 Algoritma Aho-Corasick (AC).....	11
2.3.4 Nondeterministic Finite Automaton	18
2.3.5 Pendekatan Multicharacter AC-NFA	21
2.4 Graphic Processing Unit (GPU)	26
2.4.1 The Graphics Pipeline	26
2.4.2 Evolusi Arsitektur GPU.....	27
2.4.3 GPGPU (General-Purpose Graphic Processing Unit).....	28
2.5 CUDA.....	30
BAB III ANALISIS DAN PERANCANGAN	31

3.1	Analisis	31
3.1.1	Analisis Proses	31
3.1.2	Analisis Kebutuhan Fungsional.....	36
3.1.3	Analisis Kebutuhan Non-Fungsional	44
3.2	Perancangan.....	45
3.2.1	Perancangan Tampilan	45
3.2.2	Perancangan Basis Data	58
BAB IV HASIL DAN IMPLEMENTASI		61
4.1	Hasil.....	61
4.1.1	Front-End.....	61
4.1.2	Back-End	66
4.2	Implementasi	70
4.2.1	Pengujian terhadap Sistem	70
BAB V KESIMPULAN DAN SARAN		71
5.1	Kesimpulan.....	71
5.2	Saran	71
DAFTAR PUSTAKA		72
Lampiran 1. DAFTAR RIWAYAT HIDUP		76
Lampiran 2. DAFTAR RIWAYAT HIDUP		77
Lampiran 3. DAFTAR RIWAYAT HIDUP		78

UNIVERSITAS

MIKROSKIL

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Sederhana Model <i>Information Retrieval</i>	6
Gambar 2.2 Contoh TBIR.....	7
Gambar 2.3 Proses CBIR	8
Gambar 2.4 Hasil kata kunci {he}	13
Gambar 2.5 Hasil kata kunci {she}.....	13
Gambar 2.6 Hasil kata kunci {his}	14
Gambar 2.7 Hasil kata kunci {hers}	14
Gambar 2.8 Hasil loop di <i>state</i> 0.....	15
Gambar 2.9 Fungsi <i>failure</i> untuk f(2), f(6), dan f(4)	16
Gambar 2.10 Fungsi <i>failure</i> untuk f(8) dan f(7)	16
Gambar 2.11 Fungsi <i>failure</i> untuk f(5)	17
Gambar 2.12 Fungsi <i>failure</i> untuk f(9)	17
Gambar 2.13 Hasil dari pencocokan <i>string</i>	18
Gambar 2.14 Nondeterministic Automaton N	19
Gambar 2.15 Asisten transisi	23
Gambar 2.16 $\delta^3, ??h = 1$	23
Gambar 2.17 $\delta^3, ?he = 2$	24
Gambar 2.18 $\delta^3(1, ers) = 9$ dan $\delta^3(1, e??) = -$	25
Gambar 2.19 AC-NFA dengan 9 transisi 1-karakter dan 4 <i>Output states</i>	25
Gambar 3.1 Proses Pencarian Dokumen Arsip.....	31
Gambar 3.2 Contoh fungsi <i>goto</i>	34
Gambar 3.3 3-karakter AC-NFA.....	35
Gambar 3.4 <i>Use Case</i> Sistem Pencarian Dokumen Arsip	37
Gambar 3.5 Tampilan halaman <i>login</i>	46
Gambar 3.6 Tampilan halaman <i>home</i>	47
Gambar 3.7 Tampilan halaman kelola data surat.....	48
Gambar 3.8 Tampilan halaman tambah data surat.....	49
Gambar 3.9 Tampilan halaman edit arsip	50
Gambar 3.10 Tampilan halaman scan berkas	51
Gambar 3.11 Tampilan halaman data surat <i>user</i> dan kepala sekolah	52

Gambar 3.12 Tampilan halaman <i>home</i> kepala sekolah.....	53
Gambar 3.13 Tampilan halaman kelola akun <i>user</i> dan kepala sekolah	54
Gambar 3.14 Tampilan halaman <i>home admin</i>	55
Gambar 3.15 Tampilan halaman <i>admin</i> kelola <i>user</i>	56
Gambar 3.16 Tampilan halaman uji sistem.....	57
<i>Gambar 3.17 Entity Relationship Diagram (ERD)</i>	58
Gambar 3.18 <i>Class Diagram</i> Sistem Pencarian Dokumen Arsip	59
Gambar 3.19 Model fisik ERD	59
Gambar 4.1 Tampilan halaman <i>log in</i>	61
Gambar 4.2 Tampilan Halaman Home User.....	62
Gambar 4.3 Tampilan Halaman Kelola Data Surat	62
Gambar 4.4 Tampilan Halaman Tambah Data Surat	63
Gambar 4.5 Tampilan Halaman Ubah Data Surat	64
Gambar 4.6 Tampilan Halaman Scan Berkas	65
Gambar 4.7 Tampilan Halaman Cari Data Surat	65
Gambar 4.8 Tampilan Halaman Kelola Akun.....	66
Gambar 4.9 Tampilan Halaman Log In Admin	66
Gambar 4.10 Tampilan Halaman Home Admin	67
Gambar 4.11 Tampilan Halaman Cari Data Surat	68
Gambar 4.12 Tampilan Halaman Kelola Akun User.....	69
Gambar 4.13 <i>Chart</i> Hasil Pengujian	70

UNIVERSITAS
MIKROSKIL

DAFTAR TABEL

Tabel 2.1 Fungsi <i>output</i>	17
Tabel 3.1 Fungsi <i>output</i>	34
Tabel 3.2 Aturan Otorisasi untuk Setiap Aktor.....	37
Tabel 3.3 <i>Login</i>	38
Tabel 3.4 Mengelola Data Surat.....	39
Tabel 3.5 Mencari dan Melihat Data Surat	41
Tabel 3.6 Mengelola Data User	42
Tabel 3.7 Mengelola Akun.....	43
Tabel 3.8 Keterangan PIECES.....	44
Tabel 3.9 Keterangan tb_user	59
Tabel 3.10 Keterangan tb_surat	60



UNIVERSITAS
MIKROSKIL

DAFTAR LAMPIRAN

Lampiran 1. DAFTAR RIWAYAT HIDUP	76
Lampiran 2. DAFTAR RIWAYAT HIDUP	77
Lampiran 3. DAFTAR RIWAYAT HIDUP	78



UNIVERSITAS MIKROSKIL