

KOMPARASI ALGORITMA STRING MATCHING PADA POLA TEKS

TESIS



Oleh:

ACHMAD NUR SHOLEH

1011601323

**PROGRAM STUDI MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCA SARJANA
UNIVERSITAS BUDI LUHUR**

JAKARTA

2015

KOMPARASI ALGORITMA STRING MATCHING PADA POLA TEKS

TESIS

Diajukan sebagai salah satu persyaratan
Untuk mendapatkan gelar Magister Ilmu Komputer (MKOM)



Oleh:

ACHMAD NUR SHOLEH

1011601323

**PROGRAM STUDI MAGISTER ILMU KOMPUTER (MKOM)
PROGRAM PASCA SARJANA
UNIVERSITAS BUDI LUHUR**

JAKARTA

2015

LEMBAR PERNYATAAN

Saya yang bertanda tangan dibawah ini,

Nama : Achmad Nur Sholeh
NIM : 1011601323
Program Studi : Magister Ilmu Komputer
Fakultas : Ilmu Komputer

Menyatakan bahwa TESIS yang berjudul:

KOMPARASI ALGORITMA STRING MATCHING PADA POLA TEKS

1. Merupakan hasil karya tulis ilmiah sendiri dan bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik oleh pihak lain,
2. Saya ijin untuk dikelola oleh Universitas Budi Luhur sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan penuh tanggung jawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila kemudian hari pernyataan ini tidak benar.

Jakarta, 11 Agustus 2015



Achmad Nur Sholeh

LEMBAR PENGESAHAN

Nama : Achmad Nur Sholeh
NIM : 1011601323
Bidang Peminatan : Rekayasa Komputasi Terapan
Jenjang Studi : Strata 2
Judul : KOMPARASI ALGORITMA STRING MATCHING
PADA POLA TEKS

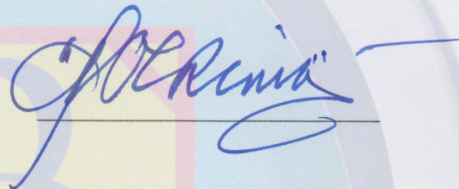
Jakarta, 11 Agustus 2015

Tim Penguji

Tanda tangan:

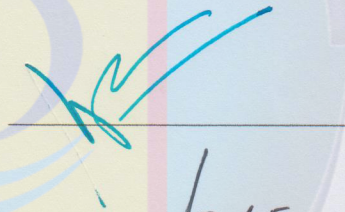
Ketua.

Prof. Dr. Moedjiono, M.Sc



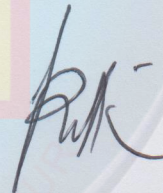
Anggota,

Ir. Wendi Usino, MM., M.Sc., Ph.D

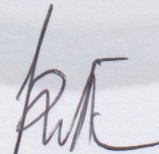


Pembimbing utama,

Dr. Ir. Nazori AZ, MT.



Ketua Program Studi



Dr. Ir. Nazori AZ, MT.

LEMBAR PERSEMBAHAN

Saya persembahkan Tesis ini untuk kedua orang tua yang selalu memotivasi penulis untuk menyelesaikan pendidikan strata dua (S2), istri dan anak-anakku tercinta Endah Wahyuningsih, Husna Hafizah Salsabila Achmad, dan Shakila Najma Ramadhani Achmad semoga kalian semua dilindungi Allah SWT.

يَا أَيُّهَا الَّذِينَ ءَامَنُوا إِن مِّنْ أَرْوَاحِكُمْ وَأَوْلَادِكُمْ عَدُوًّا
لَّكُمْ فَاحْذَرُوهُمْ وَإِن تَعَفَّوْا وَتَصَفَحُوا وَتَغْفِرُوا فَإِنَّ
اللَّهَ غَفُورٌ رَّحِيمٌ ﴿١٤﴾

Hai orang-orang mukmin, sesungguhnya di antara isteri-isterimu dan anak-anakmu ada yang menjadi musuh bagimu maka berhati-hatilah kamu terhadap mereka dan jika kamu memaafkan dan tidak memarahi serta mengampuni (mereka) maka sesungguhnya Allah Maha Pengampun lagi Maha Penyayang. (Q.S. At-Taghābun: 14)

ABSTRAK

String matching merupakan salah satu algoritma pencocokan kata (*text processing*). Menemukan *pattern* teks dalam suatu informasi yang dalam hal ini berupa abstraksi karya ilmiah mahasiswa perguruan tinggi menjadi sulit dilakukan jika terdiri dari banyak kata atau dengan kata lain semakin banyak kata menjadi semakin sulit dan memakan waktu. Implementasi algoritma *string matching* pada prototipe aplikasi deteksi kemiripan pola teks menggunakan bahasa pemrograman Ruby versi 2.1.1 dan framework Rails versi 4.1.1. Metode penelitian yang digunakan yaitu metode rekayasa menggunakan model pengembangan sistem *Rapid Application Development* yang terdiri dari tiga fase yaitu, *requirement planning*, *workshop design*, dan *implementation*. proses analisa kebutuhan menggunakan metode UML (*unified modeling language*). Penelitian ini bertujuan untuk mencari algoritma *string matching* yang efektif dengan cara membandingkan dua algoritma *string matching* yaitu *Rabin Karp* dan *Aho-Corasick*. Hasil berupa nilai persentase kemiripan pola teks (Akurasi) dan kecepatan waktu proses (*time complexity*) dari masing-masing algoritma untuk dijadikan sebagai acuan pengambilan kebijakan.

Kata kunci : *String Matching, Text Processing, Pattern, Kemiripan Pola Teks, Rabin Karp, Aho-Corasick, Time Complexity*

ABSTRACT

String matching is one word matching algorithm (text processing). Find a text pattern into an information abstraction in this case a college student scientific work becomes hard to do if it consists of many words, or in other words, the more words it becomes increasingly difficult and time consuming. String matching algorithm implementation on a prototype text pattern similarity detection applications using the Ruby programming language and Rails framework version 4.1.1. The method used is the engineering method using a model system development Rapid Application Development which consists of three phases, namely, requirements planning, workshop design, and implementation. needs analysis process using UML (unified modeling language). This study aims to find effective string matching algorithm by comparing two string matching algorithms that Rabin-Karp and Aho Corasick. Results in the form of a text pattern similarity percentage value (Accuracy) and the speed of processing time (time complexity) of each algorithm to be used as reference for policy making.

Keywords : *Matching String, Text Processing, Pattern, Pattern Similarity Text, Rabin Karp, Aho-Corasick, Time Complexity*

KATA PENGANTAR

Assalamu'alaikum, W. W.

Alhamdulillah, Puji syukur kehadiran Allah SWT yang telah memberikan berbagai macam nikmat diantaranya nikmat iman, islam serta nikmat sehat panjang umur sehingga penulisan tesis ini dapat terselesaikan. Shalawat serta salam selalu tercurah kepada Nabi Muhammad SAW. Tesis ini disusun sebagai salah satu persyaratan akademik dalam menyelesaikan jenjang pendidikan Strata 2 (S2) Program Pascasarjana Magister Ilmu Komputer Universitas Budi Luhur Jakarta.

Penyusunan tesis ini tidak dapat terlaksana dengan baik tanpa dorongan dan dukungan dari berbagai pihak yang terkait. Sebagai ungkapan rasa syukur penulis ingin mengucapkan banyak terima kasih kepada :

1. Bapak Prof. Dr. Moedjiono, M. Sc., selaku Direktur Program Pascasarjana Universitas Budi Luhur;
2. Bapak Dr. Ir. Nazori, AZ., MT., selaku Ketua Program Studi Magister Ilmu Komputer Universitas Budi Luhur dan Dosen Pembimbing, yang telah memberikan bimbingan, waktu dan perhatiannya;
3. Bapak/Ibu Dosen Magister Ilmu Komputer Universitas Budi Luhur;
4. Staff IT dan Otomasi Pusat Perpustakaan UIN Syarif Hidayatullah Jakarta, dan teman-teman Pustakawan yang tidak dapat disebutkan satu persatu;
5. Dr. Aang Subiyakto, M.Kom terima kasih telah memberikan pencerahan dalam menyelesaikan tesis ini;
6. Rekan-rekan kuliah angkatan 2011 MKOM XB, spesial thanks buat Wido Hanggoro, M.Kom atas bantuan dan motivasinya;
7. Orang tua dan keluarga yang tiada hentinya memberikan spirit, doa, dukungan, pengalaman, dan motivasi;
8. Endah Wahyuningsih istriku, Husna Hafizah Salsabila Achmad, Shakila Najma Ramadhani Achmad, kalian menjadi spirit dan semangat;
9. Pihak-pihak yang tak penulis sebutkan satu persatu. Semoga Allah SWT senantiasa membalas amal baik kita semua;

Penulis menyadari penulisan tesis ini masih jauh dari sempurna oleh karena itu penulis terbuka terhadap segala kritik dan saran yang membangun. Akhir kata penulis mempersembahkannya dengan segala kelebihan dan kekurangan, semoga penulisan tesis ini dapat bermanfaat bagi kita semua, amien.

Wassalamu 'alaikum, W. W.

Jakarta, 11 Agustus 2015

Achmad Nur Sholeh



DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERNYATAAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERSEMBAHAN	iv
ABSTRAK	v
ABSTRACT.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN.....	xiii
 BAB I PENDAHULUAN	 1
1.1 Latar Belakang	1
1.2 Masalah Penelitian	2
1.2.1 Identifikasi Masalah	2
1.2.2 Batasan Masalah	2
1.2.3 Rumusan Masalah	3
1.3 Tujuan dan Manfaat Penelitian	3
1.3.1 Tujuan	3
1.3.2 Manfaat	4
1.4 Sistematika Penulisan.....	4
1.5 Daftar Pengertian.....	5
 BAB II LANDASAN PEMIKIRAN	 6
2.1 Tinjauan Pustaka	6
2.1.1 Algoritma	6
2.1.2 <i>String Matching</i>	8
2.1.3 <i>Algoritma Rabin Karp</i>	11
2.1.4 <i>Algoritma Aho-Corasick</i>	13
2.1.5 <i>Ruby</i>	17
2.1.6 <i>Rapid Application Development (RAD)</i>	18
2.2 Tinjauan Studi	21
2.2.1 <i>The Efficiency Of String Matching Algorithms On Natural Languages Text</i>	21
2.2.2 Perbandingan Pendekatan Deteksi Plagiarism Dokumen Dalam Bahasa Inggris.....	22
2.2.3 Perancangan Algoritma Pencocokan <i>String</i> Menggunakan Fungsi <i>Hash</i> Untuk Pendeteksian Plagiarisme	22
2.2.4 Analisis Penerapan <i>String Matching</i> Pada Aplikasi Pencarian Berkas Di Komputer	22
2.2.5 Perbandingan Algoritma <i>String Matching Knuth Morris-Pratt</i> Dengan <i>Boyer Moore Horspool</i> Pada Simulasi Pengenalan Sidik Jari ..	23

2.2.6 Perbandingan Penggunaan Algoritma <i>Boyer Moore</i> dan Algoritma <i>Horspool</i> pada Pencarian <i>String</i> dalam Bahasa Medis.....	23
2.2.7 Penggunaan Algoritma <i>Apostolico-Crochemore</i> Pada Proses Pencarian <i>String</i> Di Dalam Teks.....	24
2.3 Tinjauan Objek Penelitian	28
2.3.1 Visi Pusat Perpustakaan UIN	28
2.3.2 Misi Pusat Perpustakaan UIN	28
2.3.3 Tujuan Pusat Perpustakaan UIN	29
2.3.4 Struktur Organisasi	30
2.3.5 Fakultas dan Program Studi	30
2.4 Kerangka Pemikiran	33
2.5 Hipotesis	34
BAB III DESAIN PENELITIAN	35
3.1 Metode Penelitian	35
3.2 Metode Pengumpulan Data	36
3.2.1 Studi Pustaka.....	36
3.2.2 Wawancara.....	37
3.3 Pemilihan Sampel.....	37
3.4 Instrumentasi	37
3.5 Langkah-Langkah Penelitian.....	38
3.6 Jadwal Penelitian	40
BAB IV ANALISIS, INTERPRETASI, IMPLIKASI PENELITIAN DAN RENCANA IMPLEMENTASI.....	41
4.1 Analisis Data	41
4.2 Analisis Sistem	43
4.2.1 Requirement Planning (Perencanaan Kebutuhan)	43
4.2.2 Workshop Design (Proses Desain)	44
4.2.3 Implementasi Sistem	54
4.2.4 Infrastruktur Jaringan	56
4.2.5 Implementasi Program	56
4.3 Hasil Penelitian.....	65
4.4 Implikasi Penelitian	70
4.4.1 Aspek Sistem	70
4.4.2 Aspek Manajerial	71
4.4.3 Aspek Penelitian Lanjut	71
4.5 Rencana Implementasi	71
BAB V PENUTUP	73
5.1 Kesimpulan	73
5.2 Saran	73
DAFTAR PUSTAKA	74
LAMPIRAN	77
RIWAYAT HIDUP	113

DAFTAR GAMBAR

Gambar	Halaman
II.1. Pencocokan String.....	9
II.2. Contoh Keyword Trie	14
II.3. Keyword Trie Dengan Penambahan Fungsi-Fungsi Transisi	16
II.4. Siklus Rapid Application Development.....	19
II.5. Struktur Organisasi Pusat Perpustakaan.....	30
II.6. Kerangka Pemikiran.....	33
III.1. Rapid Application Development (RAD)	35
III.2. Langkah-Langkah Penelitian	38
IV.1. Abstrak Dalam Bahasa Inggris Dan Arab	41
IV.2. Flowchart Alur Kerja Sistem	42
IV.3. Use Case Diagram	44
IV.4. Activity Diagram User	49
IV.5. Sequence Diagram User	51
IV.6. Deployment Diagram	51
IV.7. Alur Proses Rabin Karp	52
IV.8. Alur Proses Aho-Corasick	53
IV.9. Rancangan GUI	54
IV.10. Infrastruktur Jaringan	56
IV.11. Input Teks Abstrak	61
IV.12. N-Gram Teks Pertama dan Kedua Rabin Karp	61
IV.13. Hash Teks Pertama dan Kedua.....	62
IV.14. Fingerprint Teks Pertama dan Kedua	62
IV.15. Similarity Fingerprint	62
IV.16. N-Gram Teks Pertama dan Kedua Aho-Corasick	63
IV.17. Match Pattern Teks Pertama dan Kedua	63
IV.18. Kesamaan Pattern Teks	63
IV.19. Running Time	64
IV.20. Grafik Pengujian Black Box	65
IV.21. Hasil Komparasi Teks Abstrak	66
IV.22. Grafik Tingkat Kemiripan Bahasa Indonesia	67
IV.23. Grafik Waktu Proses Bahasa Indonesia.....	67
IV.24. Grafik Tingkat Kemiripan Bahasa Arab.....	68
IV.25. Grafik Waktu Proses Bahasa Arab	68
IV.26. Grafik Tingkat Kemiripan Bahasa Inggris	69
IV.27. Grafik Waktu Proses Bahasa Inggris.....	69

DAFTAR TABEL

Tabel	Halaman
I.1. Tabel Pengertian.....	5
II.1. Pseudocode Rabin Karp	12
II.2. Daftar Kecocokan Pola Pada Automata dan Suffix Link.....	17
II.3. Output Yang Dihasilkan Oleh State Terminal	17
II.4. Tinjauan Studi Penelitian	25
II.5. Program Studi.....	30
III.1. Spesifikasi Perangkat	37
III.2. Jadwal Penelitian	40
IV.1. Use Case Jalankan Program	45
IV.2. Use Case Masukkan N-Gram	45
IV.3. Use Case Teks Pertama	46
IV.4. Use Case Teks Kedua	46
IV.5. Use Case Kalkulasi	47
IV.6. Use Case Rabin Karp.....	47
IV.7. Use Case Aho-Corasick	48
IV.8. Use Case Running Time	48
IV.9. Use Case Re-Count	49
IV.10. Spesifikasi Perangkat Keras	55
IV.11. Spesifikasi Perangkat Lunak	56
IV.12. Komparasi Judul Karya Ilmiah	60
IV.13. Sampel Pengujian Black Box	64
IV.14. Hasil Komparasi Abstrak	66
IV.15. Rencana Implementasi	72

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Hasil Wawancara.....	77
2. Keputusan Rektor tentang Kode Etik Mahasiswa	81
3. Source Code.....	93
4. Hasil Pengujian.....	106



DAFTAR PUSTAKA

- [AGUSTINUS 2002] Agustinus, N. *Studi Analisis Rapid Application Development Sebagai Salah Satu Alternatif Metode Pengembangan Perangkat Lunak*. Jurnal Informatika, Vol.3, 2002. Hal 74-79. Universitas Kristen Petra, 2002.
- [ARLIADINDA 2005] Arliadinda. *Pencocokan String dengan Menggunakan Algoritma Rabin-Karp dan Shift-Or*, Bandung: Sekolah Tinggi Teknologi Telkom, 2005.
- [ATMOPAWIRO 2006] Atmopawiro. *Pengkajian dan Analisa Tiga Algoritma Efisien Rabin-Karp, Knuth Morris Prat, dan Aho-Corasick Dalam Pencarian Pola Pada Suatu Teks*, Bandung : Institut Teknologi Bandung.
- [BAEDLOWI 2006] Baedlowi, Nadjib dan Deka Aditia Adam. *String Matching Menggunakan Algoritma Rabin-Karp*, Bandung: Institut Teknologi Bandung, 2006.
- [BENTLEY 2004] Whitten, J.L. & Bentley, L.D. *System Analysis & Design Methods: Sixth Edition*, New York: Mc.Graw-Hill, 2004.
- [CHARRAS 1997] Christian Charras, et al. *Handbook Of Exact String-Matching Algorithms*, 1997.
- [CORASICK 1975] Aho, A.V. & Corasick, M. J.. *Efficient String Matching: An Aid to Bibliographic Search*, Communications Of The ACM. Volume 18, Issue 6, p333-340, 1975.
- [ERSIN 2007] Ersin, Abdul Kadir, et al. *The Efficiency of String Matching Algorithm on Natural Language Teks*, International Scientific Conference of Unitech, PP. 349-352, Gabrovo, 2007.
- [HANDOKO 2014] Handoko, Andrew. *File Undelete Untuk Memulihkan File Yang Telah Terhapus Dari File System Dengan Algoritma Aho-Corasick*, Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara, 2014.

- [HASTRIANA 2010] Hastriana, Endang. *Perbandingan Algoritma String Matching Knuth Morris Pratt Dengan Aho-Corasick Horspool Pada Simulasi Pengenalan Sidik Jari*, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2010.
- [KURNIAWATI 2008] Kurniawati, Ana dan I Wayan Simri Wicaksana. *Perbandingan Pendekatan Deteksi Plagiarism Dokumen Dalam Bahasa Inggris*, Proceeding, Seminar Ilmiah Nasional Komputer dan Sistem Intelijen (KOMMIT 2008), ISSN: 1411-6286, PP. 284-291, 2008.
- [KUSMANA 2010] Kusmana, Suherli. *Merancang Karya Tulis Ilmiah*, Bandung:PT Remaja Rosdakarya, 2010.
- [KOMPAS 2012] Lusiana Kus Anna. *Dugaan Plagiat PTN diselidiki*, 2012, <http://edukasi.kompas.com/read/2012/06/06/09355010/>. (Diakses tanggal 30 Januari 2013)
- [KENDALL 2008] Kendall, J.E. & Kendall, K.E.. *Analisis dan Perancangan Sistem*, Edisi 5, Jilid 1., Jakarta:Indeks, 2008.
- [KENDALL 2010] _____. *Analisis dan Perancangan Sistem*, Jakarta:Indeks, 2010.
- [LENZ 2008] Lenz, Patrick. *Simply Rails 2*, SitePoint Pty. Ltd. Cambridge, 2008.
- [MCLEOD 2002] McLeod, R. Jr.. *System Development: A Project Management Approach*, New York: Leigh Publishing LLC, 2002.
- [MOEDJIONO 2012] Moedjiono. *Pedoman Penelitian, Penyusunan dan Penilaian Tesis (V.5)*, Pascasarjana Universitas Budi Luhur, 2012.
- [MUNAWAR 2005] Munawar. *Pemodelan Visual dengan UML*, Graha Ilmu: Yogyakarta, 2005.
- [MUNIR 2001] Munir, Rinaldi. *Algoritma dan Pemrograman dalam Bahasa Pascal dan C*, Buku 1. Bandung:Informatika, 2001.
- [PRESSMAN 2002] Pressman, Roger S.. *Rekayasa Perangkat Lunak: pendekatan praktisi (Buku I)*. Yogyakarta:ANDI, 2002.
- [PANDAWA 2014] Pandawa, Jesry Prasasty. *Perancangan Sistem Deteksi Kemiripan Dokumen Menggunakan Algoritma Winnowing*, Karya Ilmiah Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2014.

- [PUTRA 2007] Putra, Dian Perdhana. *Perancangan Algoritma Pencocokan String Menggunakan Fungsi Hash Untuk Pendeteksian Plagiarism*, Sekolah Teknik Elektro dan informatika, Institut Teknologi Bandung, 2007.
- [RATRI 2007] Ratri, Sindy Gita. *Penggunaan Algoritma Apostolico-Crochemore Pada Proses Pencarian String Di Dalam Teks*, Sekolah Teknik Elektro dan informatika, Institut Teknologi Bandung, 2007.
- [RUBY ET ALL 2010] Ruby, Sam., Thomas, Dave., Heinemeier, David Hanson. *Agile Web Development with Rails (Fourth Edition)*, The Pragmatic Programers LLC, 2010.
- [RUBYONRAILS 2015] *Getting Started with Rails*. <http://guides.rubyonrails.org/> 2015. (diakses tanggal 7 juli 2015)
- [SYARONI 2004] Syaroni, Mokhamad dan Rinaldi Munir. *Pencocokan String Berdasarkan Kemiripan Ucapan (Phonetic String Matching) dalam Bahasa Inggris*, Bandung: Institut Teknologi Bandung, 2004.
- [SEEDGEWICK 1998] Bentley, J. & Sedgewick, R. *Ternary Search Trees*. Dr. Dobb's Journal, 1998.
- [SUARGA 2004] Suarga. *Algoritma dan Pemrograman*, Yogyakarta : ANDI, 2004.
- [SUPARDI 2009] Supardi. *Analisis Penerapan Algoritma String Matching Pada Aplikasi Pencarian Berkas di Komputer*, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta, 2009.
- [TAMBUN 2011] Tambun, Evlyn Dwi, *Perbandingan Penggunaan Algoritma BM dan Algoritma Horspool pada Pencarian String dalam Bahasa Medis*. Sekolah Teknik Elektro dan informatika, Institut Teknologi Bandung, 2011.
- [TIM PENYUSUN 2014] Tim Penyusun 2014. *Pedoman Akademik Program Strata 1 2014/2015*, Biro Akademik, Kemahasiswaan, dan Kerjasama, 2014.
- [VILO 2008] Vilo, J. *Multiple pattern matching*, 2008.
- [WIRAWAN 2008] Wirawan, Teddy Pandu. *Penggunaan Algoritma Rabin-Karp Dalam Pencocokan String*, Bandung : Institut Teknologi Bandung, 2008.