

BAB I

PENDAHULUAN

1.1 Latar belakang

Searching merupakan aktivitas yang biasa dilakukan oleh *user*. Pada penerapannya, pencarian kata bertujuan untuk menemukan kata kunci yang dicari di dalam suatu teks untuk mencari suatu informasi yang relevan dengan kata kunci, seperti mencari kata yang sama atau mengecek kesamaan antara dua buah teks. Untuk itu diperlukan suatu aplikasi yang dapat membantu *user* untuk melakukan pencarian dokumen *full* teks menggunakan algoritma pencocokan *string* agar dapat mengetahui *keyword* yang dicari terdapat di dalam isi dokumen jurnal PDF.

Terdapat berbagai macam algoritma string matching yang telah dikembangkan salah satunya yaitu algoritma Aho Corasick. Algoritma Aho Corasick adalah salah satu algoritma pencocokan string yang diciptakan oleh Alfred V. Aho dan Margaret J. Corasick. Untuk menemukan kejadian *multipattern* pada *string* teks, algoritma ini lebih tepat karena dapat melakukan pencocokan tepat dari pola dalam teks [1].

Keunggulan dari algoritma Aho Corasick *string matching* adalah dapat melakukan beberapa pencocokan kata dalam satu kali proses tanpa mengulang dari awal proses, serta fungsi *fail* yang dapat mencari karakter lain yang cocok, jika karakter yang dicari tidak cocok [2].

Algoritma Aho-Corasick memiliki kompleksitas waktu $O(m+n+z)$, dengan n merupakan panjang teks, m merupakan panjang karakter dari teks yang digunakan dalam pencarian, dan z merupakan jumlah output yang sesuai atau jumlah kemunculan pola [3].

Dari latar belakang diatas, akan dibuat aplikasi untuk pencarian dokumen *full* teks tipe PDF untuk pencarian *string* menggunakan algoritma Aho Corasick yang akan dibandingkan dengan Knuth Morris Pratt. Dengan adanya aplikasi ini diharapkan dapat mempermudah *user* dalam proses pencarian kata dan mengetahui perbandingan kecepatan pencarian kata pada algoritma Aho Corasick dan Knuth Morris Pratt.

1.2 Rumusan Masalah

1. Bagaimana implementasi algoritma Aho Corasick pada pencarian dokumen *full* teks tipe pdf
2. Bagaimana mengetahui pengaruh/korelasi jumlah dokumen *full* teks dengan kecepatan waktu pencarian *string* pada dokumen *full* teks
3. Bagaimana mengetahui hasil uji *performance* antara algoritma Aho Corasick dengan algoritma Knuth Morris Pratt

1.3 Tujuan

Adapun tujuan dari Tugas Akhir ini adalah mengimplementasikan algoritma Aho Corasick untuk *search* dokumen *full* teks tipe PDF.

1.4 Manfaat

Adapun manfaat dari Tugas Akhir ini, antara lain :

1. Membantu *user* dalam melakukan pencarian *string* sesuai *keyword* yang dicari pada dokumen *full* teks.
2. Mengetahui alur proses algoritma Aho Corasick yang diterapkan pada pencarian dokumen *full* teks.

3. Mengetahui kemampuan sistem dalam melakukan pencarian *string* pada dokumen *full* teks dilihat berdasarkan waktu dan banyaknya dokumen *full* teks yang ada.

1.5 Ruang Lingkup

Agar pengerjaan tugas akhir ini menjadi lebih terarah dan mendapatkan hasil yang lebih spesifik, maka sistem yang dirancang dibatasi pada ruang lingkup pembahasan sebagai berikut :

1. Algoritma yang digunakan dalam penelitian ini yaitu Algoritma Aho Corasick dan Algoritma Knuth Morris Pratt sebagai pembandingan.
2. File yang diproses adalah file dengan ekstensi pdf.
3. Pencarian dilakukan pada dokumen yang tersimpan di database.
4. Inputan *keyword* berupa abjad ['a...z'] kecuali spasi yang digunakan untuk membedakan kata satu dengan kata inputan lain [4].
5. Pencarian *string* dibatasi sebanyak 3 kata [5].
6. Aplikasi *search* dokumen *full* teks merupakan aplikasi berbasis web.

1.6 Metodologi Penelitian

Metodologi penelitian yang digunakan dalam tugas akhir ini antara lain :

1. Studi Literatur

Mencari dan mempelajari referensi-referensi yang berkaitan dengan algoritma Aho Corasick dan Knuth Morris Pratt.

2. Analisa

Analisa meliputi analisa terhadap pencarian *string* yaitu pencarian *string* terhadap suatu dokumen *full* teks. Selain itu juga diperlukan analisa kebutuhan untuk aplikasi pencarian.

3. Desain

Model yang digunakan untuk desain aplikasi pencarian ini adalah *UML*.

4. Implementasi

Pada tahap ini algoritma Aho Corasick diimplementasikan pada aplikasi pencarian *string* pada *search engine* dengan acuan *output* dari tahap analisa dan desain yang telah dilakukan sebelumnya.

5. Pengujian dan evaluasi

Pada tahap ini dilakukan pengujian dan evaluasi terhadap aplikasi pencarian *string* dengan algoritma Aho Corasick yang dibandingkan dengan algoritma Knuth Morris Pratt.

1.7 Sistematika Penulisan

Secara garis besar penulisan laporan tugas akhir ini terbagi dalam beberapa bab yang tersusun sebagai berikut :

BAB I PENDAHULUAN

Dalam bab I ini berisikan latar belakang, rumusan masalah, tujuan, manfaat, ruang lingkup, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada Bab II membahas tentang teori-teori yang berhubungan dengan judul tugas akhir, seperti Aho Corasick, Knuth Morriss Pratt, *String Matching* dan Uml.

BAB III ANALISIS DAN PERANCANGAN

Pada Bab III menjelaskan mengenai analisis dan perancangan tentang implementasi Aho-Corasick untuk pencarian *document full* teks tipe PDF.

BAB IV IMPLEMENTASI SISTEM DAN HASIL PENGUJIAN

Pada bab IV menunjukkan hasil implementasi dan hasil pengujian dari algoritma Aho Corasick dan algoritma Knuth Morris Pratt.

BAB V KESIMPULAN DAN SARAN

Pada Bab V berisikan kesimpulan dan saran dari hasil Tugas Akhir yang dilaksanakan.