

Penerapan Algoritma Approximate String Matching Untuk Pencarian Teks Pada Aplikasi Ensiklopedia Teknologi Komputer

Aria Novitra

Program Studi Teknik Informatika, Fakultas Ilmu Komputer & Teknologi Informasi, Universitas Budi Darma, Medan, Indonesia
Jl. Sisingamangaraja No.338, Siti Rejo I, Kec. Medan Kota, Kota Medan, Sumatera Utara, Indonesia
Email: profjonder@gmail.com

Abstrak—Teknologi komputer adalah suatu yang diciptakan untuk kepentingan dalam pengolahan data sehingga teknologi yang dimaksud adalah perkembangan yang mana suatu sistem terdahulu digantikan dengan sistem baru sehingga memudahkan user dalam menggunakannya. Manfaat komputer saat ini cukup beragam mulai sebagai alat bantu nulis, menggambar, mengedit foto, memutar video sampai analisis data hasil penelitian maupun mengoperasikan program-program. Pada penelitian ini penulis menggunakan Algoritma Approximate string matching untuk pencarian teks pada aplikasi ensiklopedia teknologi komputer. Algoritma Approximate string matching merupakan salah satu metode yang banyak diterapkan untuk mengidentifikasi dalam pencarian teks. Approximate string matching dapat digunakan untuk pencarian string berdasarkan string yang sama. Algoritma Hamming Distance adalah salah satu algoritma approximate string matching. Algoritma ini digunakan untuk teori informasi sebagai estimasi error. Pada masa sekarang algoritma ini banyak digunakan untuk teori informasi, teori pengkodean dan kriptografi. Cara kerja Algoritma Hamming Distance yaitu dengan mengukur jarak antara dua string yang ukurannya sama dengan membandingkan simbol-simbol yang terdapat pada kedua string pada posisi yang sama.

Kata Kunci: Ensiklopedia, Algoritma Hamming Distance, String Matching, Pencarian Teks

Abstract—Computer technology is something that was created for the benefit of data processing so that the technology in question is a development in which an old system is replaced with a new system so that it makes it easier for users to use it. The benefits of computers today are quite diverse, starting as a tool for writing, drawing, editing photos, playing videos, to analyzing research data and operating programs. In this study the authors used the Approximate string matching algorithm to search text in computer technology encyclopedia applications. Approximate string matching algorithm is one method that is widely applied to identify in text search. Approximate string matching can be used to search for strings based on the same string. The Hamming Distance algorithm is one of the approximate string matching algorithms. This algorithm is used for information theory as error estimation. At present this algorithm is widely used for information theory, coding theory and cryptography. The way the Hamming Distance Algorithm works is by measuring the distance between two strings of the same size by comparing the symbols contained in the two strings in the same position.

Keywords: Encyclopedia, Hamming Distance Algorithm, String Matching, Text Search

1. PENDAHULUAN

Ensiklopedia adalah karya referensi atau abstrak yang memberikan ringkasan informasi dari semua cabang ilmu atau bidang tertentu. Ensiklopedia dibagi menjadi artikel atau entri [1]. Entri ini biasanya disusun dalam urutan abjad dan terkadang diklasifikasikan berdasarkan kategori populer. Menurut kamus besar bahasa Indonesia ensiklopedia adalah buku (serangkaian buku) yang menghimpun keterangan atau uraian tentang berbagai hal dalam bidang seni dan ilmu pengetahuan yang di susun menurut abjad atau lingkungan ilmu. Sedangkan pengertian ensiklopedia menurut Abdul Chear adalah jenis kamus yang selain memberikan keterangan makna kata, juga memuat keterangan tentang sesuatu [2]

Ensiklopedia telah ada selama sekitar 2.000 tahun yang lalu dan telah berevolusi sejak saat itu hingga bahasa (ditulis dalam bahasa internasional atau bahasa vernakular), ukuran (beberapa atau banyak volume), maksud (presentasi dari pengetahuan global atau jangkauan pengetahuan terbatas), persepsi budaya (latar belakang, minat, kelayakan, kemampuan), dan distribusi (manuskrip tulisan tangan, kecil atau besar), persepsi budaya. Sebagai sumber terpercaya yang dikumpulkan oleh para ahli, versi cetak banyak di temukan di perpustakaan, sekolah dan institusin pendidikan lainnya. Munculnya versi digital dan open source di abad ke – 20 telah memperluas jangkauan aksesibilitas, ke pengurangan, pembaca, dan Variasi entri ensiklopedia dan mempertanyakan gagasan tentang ensiklopedia dan relevansi penerapan pada produksi dinamis seperti tradisional.

Ensiklopedia Digital/Aplikasi Ensiklopedia adalah media yang Berupa teks, grafik, audio atau animasi yang berisi penjelasan komperensif mengenai keseluruhan cabang ilmu pengetahuan. Ensiklopedia biasanya menjelaskan secara khusus satu cabang ilmu pengetahuan tertentu. Ensiklopedia ini tersusun dalam bagian-bagian artikel-artikel dengan satu topik bahasa.Untuk mempermudah para pengguna dalam pencarian teks pada aplikasi Ensiklopedia dalam permasalahan ini penulis ingin menerapkan Algoritma *String Matching* dan *Algoritma Approximate*. *String matching* adalah pencarian sebuah *pattern* pada sebuah teks [3].

String matching digunakan untuk menemukan suatu *string* yang disebut dengan *pattern* dalam *string* yang disebut dengan teks. Algoritma *string matching* mempunyai 2 komponen utama, yaitu: *Pattern* dan *Text*. Algoritma *string matching* yang digunakan untuk pencarian teks pada aplikasi Ensiklopedia pada penelitian ini adalah Algoritma *Approximate* [3]-[4].

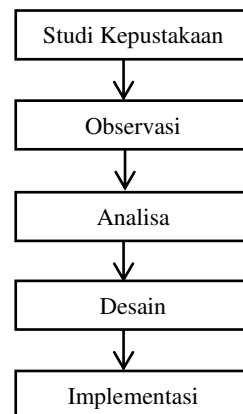
Algoritma *Approximate string matching* merupakan salah satu metode yang banyak di terapkan untuk mengidentifikasi dalam pencarian teks. *Approximate string matching* dapat di gunakan untuk pencarian string

berdasarkan string yang sama. Algoritma Hamming Distance adalah salah satu algoritma approximate string matching. Algoritma ini digunakan untuk teori informasi sebagai estimasi Error. Pada masa sekarang algoritma ini banyak digunakan untuk teori informasi, teori pengkodean dan kriptografi. Cara kerja Algoritma Hamming Distance yaitu dengan mengukur jarak antara dua string yang ukurannya sama dengan membandingkan simbol-simbol yang terdapat pada kedua string pada posisi yang sama [5]-[4].

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Adapun tahapan-tahapan yang dilakukan pada penelitian ini dapat dijelaskan seperti gambar di bawah ini:



Gambar 1. Tahapan Penelitian

Berikut penjelasan gambar 1 diatas mengenai tahapan penelitian yang digunakan dalam penelitian :

- Studi Kepustakaan (*Library*)**
Melakukan studi kepustakaan bertujuan untuk mendapatkan referensi atau sumber informasi mengenai algoritma *approximate* yang digunakan dan aplikasi ensiklopedia yang dibangun pada penelitian ini.
- Observasi**
Metode ini dilakukan dengan melakukan pengamatan terhadap sistem pencarian yang sudah ada.
- Analisa (*Analysis*)**
Menganalisa permasalahan dan mempelajari prosedur pencocokan *string* menggunakan algoritma *Approximate*.
- Desain (*Design*)**
Mendesain prosedur sistem, Mendesain tampilan aplikasi ensiklopedia, dan Implementasi Algoritma *Approximate*.
- Implementasi (*Implementation*)**
Mengimplementasikan Aplikasi Ensiklopedia yang telah dibangun pada *smartphone android* yang didukung sistem operasi *android versi 4.0* sampai *android versi 6.0*.

2.2 Ensiklopedia

Ensiklopedia merupakan karya referensi, ringkasan, abstrak yang memberikan informasi dari semua cabang ilmu dibidang tertentu. Ensiklopedia dibagi menjadi artikel atau entri, biasanya dalam urutan abjad, terkadang menurut kategori subjek [3]-[6].

2.3 Algoritma *Hamming Distance*

Algoritma "hamming"distance" merupakan metode untuk mengukur jarak antara dua string yang ukurannya sama dengan membandingkan simbol-simbol yang terdapat pada kedua string pada posisi yang sama. Hamming distance dari dua string adalah jumlah simbol dari kedua string yang berbeda. Contoh : Hamming distance antara string 'toned' dan 'roses' adalah 3. Hamming Distance digunakan untuk mengukur jarak antar dua string binary misalnya jarak antara 10011101 dengan 10001001 adalah 2. Algoritma Haming distance dipakai untuk mencari seberapa mirip sebuah vektor terhadap vektor yang lain berdasarkan nilai kedekatannya. Jika nilai kedekatan semakin kecil maka kemiripan kedua vektor semakin besar sebaliknya jika nilai kedekatan semakin besar kemiripan kedua vektor semakin kecil [3]-[7].

2.4 Algoritma String Matching

Algoritma *String Matching* adalah algoritma yang dapat digunakan untuk mencari *string* pendek (pola) yang ada di dalam teks. Contoh penerapan pencocokkan string adalah pencocokkan string di Microsoft word, editor, atau pencocokkan situs web dengan memasukkan kata kunci yang telah diterapkan ke mesin pencari seperti yahoo dan google [8]-[9]. Pembagian algoritma string matching berdasarkan arah pencocokan stringnya adalah sebagai berikut :

- Arah pembacaan dari kiri ke kanan.

Algoritma yang termasuk kategori ini adalah *Brute Force*, *Levenshtein Distance*, *Morris* dan *Pratt* (yang kemudian dikembangkan oleh Knuth, Morris, dan Pratt).

- b. Arah pembacaan dari kanan ke kiri.

Algoritma yang termasuk kategori ini adalah Boyer dan Moore yang kemudian dikembangkan menjadi algoritma *turbo Boyer-Moore*, *tuned Boyer-Moore*, dan *Zhu-Takaoka*. Arah pencarian yang ditentukan program. Algoritma yang termasuk kategori ini adalah algoritma *Colussi*, *Crochemore-Perrin*.

2.5 pencarian Teks

Dalam ilmu komputer, salah satu algoritma pencarian di jelaskan secara luas adalah salah satu algoritma yang menerima masukkan berupa sebuah masalah dan menghasilkan sebuah solusi untuk masalah tersebut, yang biasanya didapat dari evaluasi beberapa kemungkinan solusi. Himpunan semua kemungkinan solusi dari sebuah masalah di sebut ruang pencarian [10]-[11]. Algoritma pencarian brute-force atau pencarian naif/ uninformed memakai metode yang sederhana dan sangat intuitif pada ruang pencarian, sedangkan algoritma pencarian informed menggunakan heuristik untuk menerapkan pengetahuan tentang struktur dari ruang pencarian untuk berusaha mengurangi banyaknya waktu yang dipakai dalam pencarian. Teks adalah susunan kata yang digunakan untuk memberikan informasi, menjelaskan arti, dll [12].

3. HASIL DAN PEMBAHASAN

3.1 Analisa

Analisa ini akan mendiskripsikan bagaimana proses sistem yang akan dibangun untuk menyelesaikan permasalahan yang diangkat di dalam penelitian ini. Dalam hal ini, peneliti menggunakan algoritma *Hamming Distance* untuk pencarian teks pada aplikasi ensiklopedia teknologi komputer.

Penerapan algoritma *Hamming Distance* adalah cara untuk mengukur jarak antara 2 string yang ukurannya sama dengan membandingkan simbol-simbol yang terdapat pada 2 string pada posisi yang sama. Dalam pencarian teks *Hamming distance* digunakan untuk memperbaiki telekomunikasi sebagai esimasi error. *Hamming distance* dari dua string adalah jumlah simbol dari kedua string yang berbeda.

Dalam melakukan suatu perancangan maka harus dilakukan sesuai dengan tahapan aplikasi yang akan dibangun. Dengan adanya kesalahan pada teks maka akan lebih mudah membangun aplikasi yang telah dirancang dengan algoritma yang telah ditentukan. Tujuan tahap analisa ini adalah sebagai aktivitas teknis yang mengidentifikasi data, pencarian teks dan kebutuhan perancangan aplikasi untuk pencarian teks pada aplikasi ensiklopedia teknologi komputer. Adapun kegiatan yang dilakukan diantaranya adalah:

- Menganalisis kebutuhan sistem yang akan dibuat.
- Menganalisis aspek yang mendukung keberhasilan sistem.
- Menganalisis *stakeholder* yang terlibat dalam sistem.

3.1.2 Penerapan Algoritma Hamming Distance

Berdasarkan penelitian, penggunaan algoritma jarak hamming merupakan metode untuk menghitung jumlah selisih dua file biner dengan panjang yang sama. Algoritma jarak hamming membantu pencarian teks dengan mendeteksi dan mengoreksi perkiraan kesalahan telekomunikasi. Algoritma ini digunakan untuk teori informasi sebagai estimasi error. Cara kerja *Algoritma Hamming Distance* yaitu dengan mengukur jarak antara dua string yang ukurannya sama dengan membandingkan simbol-simbol yang terdapat pada kedua string pada posisi yang sama.

Contoh penggunaan algoritma hamming distance untuk pencarian pattern dalam teks:

Jika diberikan : $x = \text{sate} \rightarrow |x| = n = 4$

$y = \text{soto} \rightarrow |y| = n = 4$

maka, perjalanan algoritmanya adalah sebagai berikut:

$$n = 4 > 0.$$

$$C_{0,0} = 0.$$

$i = 1$

$j = 1, i = j, x_1 = s = y_1$, maka $C_{1,1} = C_{0,0} + 0 = 0$.

$j = 2, i \neq j$, maka $C_{1,2} = 0$. $j = 3, i \neq j$, maka $C_{1,3} = 0$.

$j = 4, i \neq j$, maka $C_{1,4} = 0$.

$i = 2$.

$j = 1, i \neq j$, maka $C_{2,1} = 0$.

$j = 2, i = j, x_2 = a \neq o = y_2$, maka $C_{2,2} = C_{1,1} + 1 = 1$.

$j = 3, i \neq j$, maka $C_{2,3} = 0$.

$j = 4, i \neq j$, maka $C_{2,4} = 0$.

$i = 3$.

$j = 1, i \neq j$, maka $C_{3,1} = 0$.

$j = 2, i \neq j$, maka $C_{3,2} = 0$.

$j = 3, i = j, x_3 = t = y_3$, maka $C_{3,3} = C_{2,2} + 0 = 1$.
 $j = 4, i \neq j$, maka $C_{3,4} = 0$.
 $i = 4 = n$.
 $j = 1, i \neq j$, maka $C_{4,1} = 0$.
 $j = 2, i \neq j$, maka $C_{4,2} = 0$.
 $j = 3, i = j$, maka $C_{4,3} = 0$.
 $j = 4 = n, i \neq j, x_4 = e \neq o = y_4$, maka $C_{4,4} = C_{3,3} + 1 = 2$.
 Output : 2. Artinya, jarak antara x dan y adalah 2.

Tabel 1. perhitungan jarak *Hamming* pada contoh diatas dengan menggunakan tabel

-	-	0	1	2	3	4
-	-	-	s	a	t	e
0	-	0	0	0	0	0
1	s		0	0	0	0
2	o	0	0	1	0	0
3	t	0	0	0	1	0
4	e	0	0	0	0	2

Dari contoh di atas, terlihat bahwa sel $C_{nn} = C_{4,4} = 2$ artinya jarak *Hamming* antara teks x dan y adalah 2.

3.2 Implementasi

Aplikasi Ensiklopedia Teknologi Komputer yang akan dirancang pada digital yang mendukung pemograman android. Sebelum menjalankan aplikasi untuk pencarian teks pada digital yang digunakan.. Dalam pembuatan aplikasi ensiklopedia teknologi kompueter, ada beberapa hal yang penting yaitu perancangan model diagram dan perancangan antarmuka. Proses pencarian teks berdasarkan algoritma yang digunakan, termasuk dalam hal penting pembuatan aplikasi ensiklopedia teknologi komputer ini karena proses tersebut merupakan bagian inti dari penggunaan aplikasi teknologi komputer tersebut. Berikut dapat dilihat tampilan aplikasi yang sudah dirancang :

a. Menu Utama

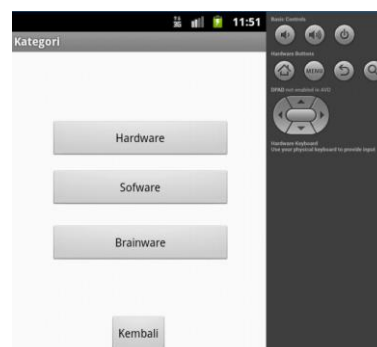
Menu utama merupakan media penghubung pengguna aplikasi dengan aplikasi Ensiklopedia Teknologi Komputer yang dibangun oleh penulis pada penelitian ini untuk memilih menu yang ada pada aplikasi Ensiklopedia Teknologi Komputer yang dibangun. Menu utama pada aplikasi Ensiklopedia Teknologi Komputer yang dibangun oleh penulis pada penelitian ini seperti gambar berikut:



Gambar 2. Menu Utama

b. Kategori

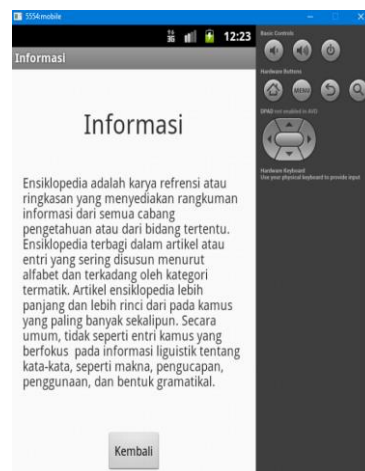
Kategori merupakan interface penghubung yang dapat digunakan pengguna untuk melihat beberapa kategori yang ingin dilihat pada aplikasi Ensiklopedia Teknologi Komputer yang dibangun oleh penulis pada penelitian ini seperti gambar berikut:



Gambar 3. Tampilan Kategori

c. Informasi

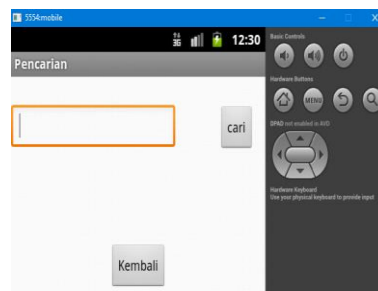
Tampilan ini merupakan interface penghubung yang berisi tentang informasi sekitar aplikasi Ensiklopedia Teknologi Komputer ini. Adapun rancangan aplikasi yang dibangun oleh penulis adalah sebagai berikut:



Gambar 4. Tampilan Informasi

d. Pencarian

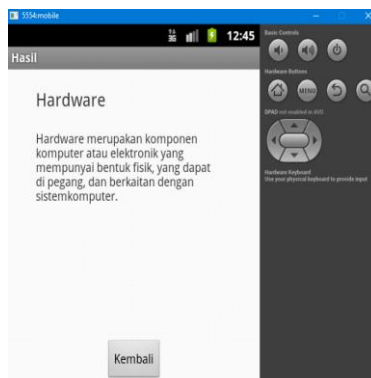
Pencarian merupakan media penghubung pengguna aplikasi dengan aplikasi Ensiklopedia Teknologi Komputer yang dibangun oleh penulis pada penelitian ini untuk melakukan proses pencarian istilah pada Ensiklopedia Teknologi Komputer. Pencarian pada aplikasi Ensiklopedia Teknologi Komputer yang dibangun oleh penulis pada penelitian ini seperti gambar berikut:



Gambar 5. Pencarian

e. Hasil Pencarian

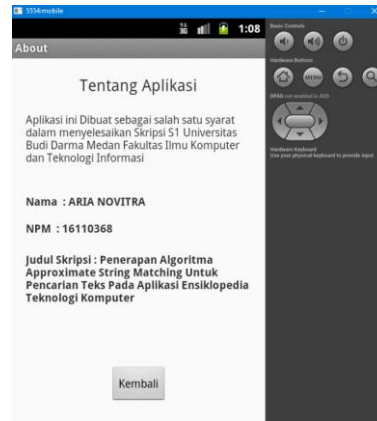
Hasil pencarian merupakan media penghubung pengguna aplikasi dengan aplikasi Ensiklopedia Teknologi Komputer yang dibangun oleh penulis pada penelitian ini untuk menampilkan hasil pencarian isi Ensiklopedia Teknologi Komputer. Hasil pencarian pada aplikasi Ensiklopedia Teknologi Komputer yang dibangun oleh penulis pada penelitian ini seperti gambar berikut:



Gambar 6. Hasil Pencarian

f. Tentang Aplikasi

Tentang aplikasi merupakan media penghubung pengguna aplikasi dengan aplikasi Ensiklopedia Teknologi Komputer yang dibangun oleh penulis pada penelitian ini untuk menampilkan informasi mengenai tentang aplikasi Ensiklopedia Teknologi Komputer ini. Adapun tampilan yang dibangun oleh penulis pada penelitian ini seperti gambar berikut:



Gambar 7. Tentang Aplikasi

4 KESIMPULAN

Berdasarkan analisis dan pembahasan dan bab sebelumnya, maka penulis dapat menarik kesimpulan bahwa dalam pencarian teks pada aplikasi ensiklopedia teknologi komputer maka dapat diterapkan dengan menggunakan algoritma hamming distance. dan dengan adanya aplikasi yang telah dibangun dapat membantu para pengguna (user) dalam pencarian kesalahan yang ada pada teks.

REFERENCE

- [1] J. Jatminto and I. K. D. Nuryana, "Implementasi Spelling Checker dengan Algoritma Levenshtein distance pada Ensiklopedia IT (Information Technology) berbasis website," *Inov. J. Ilm. Inov. Teknol. Inf.*, vol. 1, no. 1, 2016.
- [2] A. F. Siregar, "Perancang Aplikasi Ensiklopedia Budaya Indonesia Dengan Menerapkan Algoritma Crochemore-Perrin Untuk Pencarian Teks," *J. Glob. Technol. Comput.*, vol. 2, no. 1, pp. 16–23, 2022.
- [3] Y. N. Gulo, "Penerapan Algoritma Hamming Distance Untuk Pencarian Teks Pada Aplikasi Ensiklopedia Indonesia," *J. Glob. Technol. Comput.*, vol. 1, no. 2, pp. 50–54, 2022.
- [4] E. S. Nasution, N. A. Hasibuan, and S. Suginam, "Implementasi algoritma approximate string matching pada aplikasi filosofi berbasis android," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, 2018.
- [5] A. Ervana and A. Pertiwi, "Implementasi Algoritma Pencocokan String pada Aplikasi Pengarsipan Berbasis Web," *J. Inform.*, vol. 3, no. 2, pp. 1–14, 2012.
- [6] A. S. Dharma, J. Banjarnahor, O. Nainggolan, and Y. Sihombing, "Analisis Algoritma Approximate String Matching Pada Fitur Autocorrect dalam Pencarian Data," *J. Nas. Tek. Inform. dan Elektro*, vol. 1, no. 01, pp. 1–6, 2018.
- [7] B. M. D. Adiwidya, "Algoritma levenshtein dalam pendekatan approximate string matching," *Bandung Inst. Teknol. Bandung*, 2009.
- [8] Y. Rochmawati and R. Kusumaningrum, "Studi Perbandingan Algoritma Pencarian String dalam Metode Approximate String Matching untuk Identifikasi Kesalahan Pengetikan Teks," *J. Buana Inform.*, vol. 7, no. 2, 2016.
- [9] N. Fadhilah, H. Azis, and D. Lantara, "Validasi Pencarian Kata Kunci Menggunakan Algoritma Levenshtein Distance Berdasarkan Metode Approximate String Matching," in *Prosiding SAKTI (Seminar Ilmu Komputer dan Teknologi Informasi)*, 2018, vol. 3, no. 2, pp. 129–133.
- [10] R. I. Borman, "Penerapan String Matching Dengan Algoritma Boyer Moore Pada Aplikasi Font Italic Untuk Deteksi Kata Asing," *J. Teknoinfo*, vol. 10, no. 2, pp. 39–43, 2016.
- [11] E. Haryanto, "Rancang Bangun Prototype Mesin Pencari String Menggunakan Metode Fuzzy String Matching," *Konf. Nas. Sist. dan Inform. KNS&III-012*, pp. 76–82, 2011.
- [12] A. Aradea, "Aplikasi SMS Gateway dengan Koreksi Kesalahan Menggunakan Fuzzy String Matching," 2007.