

# Penerapan Algoritma Hamming Distance Untuk Pencarian Teks Pada Aplikasi Ensiklopedia Indonesia

Yunimawar Niati Gulo

Prodi Studi Teknik Informatika, Fakultas Ilmu Komputer Dan Teknologi Informasi, Universitas Budi Darma  
Jl. Sisingamangaraja No. 338, Medan, Sumatera Utara, Indonesia  
Email: [mawargilsev2910@gmail.com](mailto:mawargilsev2910@gmail.com)

**Abstrak**– String Matching (pencocokan string) adalah algoritma untuk melakukan pencarian semua kemunculan string pendek (pattern) yang muncul dalam teks. Contoh implementasi string matching adalah pencocokan string pada Microsoft word, editor atau pencocokan website dengan memasukkan kata kunci sebagai mana yang telah diterapkan pada search engine seperti yahoo, atau google. Algoritma Hamming Distance merupakan salah satu dari 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 ke dua string pada posisi yang sama. Maka dari itu pada metode atau algoritma diatas dalam Skripsi ini Penulis ingin mengimplementasikan Penerapan algoritma Hamming Distance untuk pencarian teks pada aplikasi Ensiklopedia Indonesia.

**Kata Kunci** : String Matching, Algoritma Hamming Distance, *Visual Studio 2008*.

**Abstract**– String Matching is an algorithm for searching all occurrences of short strings (patterns) that appear in the text. Examples of implementing string matching are string matching in Microsoftword, editor or web matching by entering in cyanwords as which have been applied to search engines such as yahoo, or google. Hamming distance matching algorithm is one of the algorithm for information theory as an estimate error. At present this algorithm is widely used for information theory, coding theory and cryptograpy. The Hamming Distance algorithm works by measuring the distance between two string pf the same size by comparing the symbols contained in the two string at the same position. Therefore, in the method or algorithm above in this thesis, the author wants to implement the Hamming Distance algorithm application for text search in the Indonesian Encyclopedia Application.

**Keywords**: String Matching, Hamming Distance Algorithm, *Visual Studio 2008*.

## 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. Entri atau entri ini biasanya disusun dalam urutan abjad dan terkadang diklasifikasikan berdasarkan kategori populer. Artikel ensiklopedia bahkan lebih panjang dan lebih rinci daripada kamus yang paling luas. Secara umum, tidak seperti entri kamus yang berfokus pada informasi linguistik (seperti makna, pengucapan, penggunaan, dan bentuk tata bahasa) kata, artikel ensiklopedia berfokus pada informasi faktual tentang subjek yang disebutkan dalam judul artikel.

Ensiklopedia berusia sekitar 2.000 tahun, dan sejak itu berkembang menjadi bahasa (ditulis dalam bahasa internasional atau vernakular), ukuran (beberapa atau beberapa volume), niat (presentasi pengetahuan global atau ruang lingkup pengetahuan terbatas), dan konsep budaya Konsep budaya (latar belakang, minat, kualifikasi, kemampuan) dan sebaran (naskah tulisan tangan, ukuran). Sebagai sumber terpercaya yang dikumpulkan oleh para ahli, versi cetak biasanya ditemukan di perpustakaan, sekolah, dan lembaga pendidikan lainnya.

Munculnya versi digital dan open source di abad ke-20 memperluas aksesibilitas, ruang lingkup penulis, pembaca dan ensiklopedia, serta mempertanyakan konsep ensiklopedia dan relevansinya dengan produksi dinamis (seperti tradisional). Ensiklopedia Digital/ Aplikasi Ensiklopedia adalah media yang Berupa teks, grafik, audio atau animasi yang berisi penjelasan komprehensif 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 Hamming Distance.

String matching adalah pencarian sebuah pattern pada sebuah teks. 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 Hamming Distance.

Algoritma jarak hamming merupakan salah satu algoritma pendekatan pencocokan string yang digunakan dalam teori informasi sebagai estimasi kesalahan. Saat ini, algoritma ini telah banyak digunakan dalam teori informasi, teori pengkodean dan kriptografi. Algoritma Hamming distance bekerja dengan cara membandingkan simbol-simbol yang terdapat pada dua string pada posisi yang sama untuk mengukur jarak antara dua string yang berukuran sama.

## 2. METODOLOGI PENELITIAN

### 2.1 Kerangka Kerja Penelitian

Metode Penelitian yang digunakan untuk mengumpulkan data sebagai berikut:

- a. Perpustakaan (*Library*)  
Melakukan studi kepustakaan bertujuan untuk mendapatkan referensi atau sumber informasi mengenai algoritma *Hamming Distance* yang akan di terapkan pada Aplikasi Ensiklopedia.
- b. Analisa (*Analysis*)  
Menganalisa permasalahan dan mempelajari prosedur pencocokan string menggunakan algoritma Hamming Distance.
- c. Implementasi  
Tahap ini digunakan untuk menerapkan algoritma Metode Hamming Distance untuk Pencarian Text pada Aplikasi Ensiklopedia.

## 2.2 Kecerdasan Buatan

Kecerdasan buatan adalah bagian dari ilmu komputer, yang secara khusus telah dibuktikan dalam desain otomasi perilaku dalam sistem kecerdasan komputer. Bagian utama dari kecerdasan buatan adalah basis pengetahuan, yaitu pemahaman atau pemahaman tentang bidang pelajaran yang diperoleh melalui pembelajaran dan pengalaman [1].

## 2.3 Algoritma String Matching

Algoritma pencocokan string merupakan algoritma yang digunakan untuk mencari string pendek (pola) yang muncul dalam teks. Polanya berupa string  $m$  karakter ( $m < n$ ). Teks adalah panjang string, dan panjangnya adalah  $n$  karakter. Contoh penerapan pencocokan string adalah pencocokan string di Microsoft word, editor, atau pencocokan situs web dengan memasukkan kata kunci yang telah diterapkan ke mesin pencari seperti yahoo atau google. Pembagian algoritma pencocokan string berdasarkan arah pencocokan string adalah sebagai berikut [2]:

- a. 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 pembaca dari kanan ke kiri.  
Algoritma yang termasuk kategori ini adalah Boyer dan Moore yang kemudian di kembangkan menjadi algoitma turbo Boyer - Moore, tuned Boyer - Moore, dan Zhu - Takaoka.
- c. Arah pencarian yang ditentukan program. Algoritma yang termasuk kategori ini adalah algoritma Colussi, Crochemore-Perrin.

## 2.4 Algoritma Hamming Distance

Algoritma Hamming Distance merupakan salah satu dari algoritma Approximate String yang ditemukan oleh Richar Hamming, pada tahun 1950. Algoritma Hamming distance dapat digunakan untuk menghitung jumlah perbedaan dua buah biner yang mempunyai panjang yang sama[3].

## 2.5 Ensiklopedia

Ensiklopedia adalah karya referensi atau abstrak yang memberikan ringkasan informasi dari semua cabang ilmu atau bidang tertentu. Ensiklopedia dibagi menjadi artikel atau entri, biasanya dalam urutan abjad, terkadang menurut kategori subjek[7].

# 3. HASIL DAN PEMBAHASAN

Penerapan algoritma jarak hamming merupakan metode pengukuran jarak antara dua string yang berukuran sama dengan membandingkan simbol-simbol yang terdapat pada dua string pada posisi yang sama. Dalam penelusuran teks, jarak Hamming digunakan untuk meningkatkan estimasi telekomunikasi yang salah. Jarak Hamming dari dua string adalah jumlah simbol dalam dua cincin berbeda.

## 3.1 Penerapan Algoritma Hamming Distance

Berdasarkan penelitian, penggunaan algoritma jarak hamming merupakan metode untuk menghitung jumlah selisih dua file biner dengan panjang yang sama. Algoritme jarak Hamming membantu pencarian teks dengan mendeteksi dan mengoreksi perkiraan kesalahan telekomunikasi. 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.

Adapun tahap dalam melakukan pencarian teks agar dapat mempermudah pencarian teks adalah sebagai berikut :

Jika diberikan :  $x = \text{sate}$

$\rightarrow |x| = n = 4$

$y = \text{soto}$

$\rightarrow |y| = n = 4$

maka, perjalanan algoritmanya adalah sebagai berikut:

$n = 4 > 0$ .

$i = 1$ .

$j = 1, i = j, x_i = s = y_1$ , 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  $C_{4,3} = 0$ .

$j = 4, i = j, X_4 = e \neq o = y_4$ , maka  $C_{4,4} = C_{3,3} + 1 = 2$ .

Output : 2. Artinya, jarak antara  $x$  dan  $y$  adalah 2.

Tabel perhitungan jarak *Hamming* pada contoh diatas dengan menggunakan tabel:

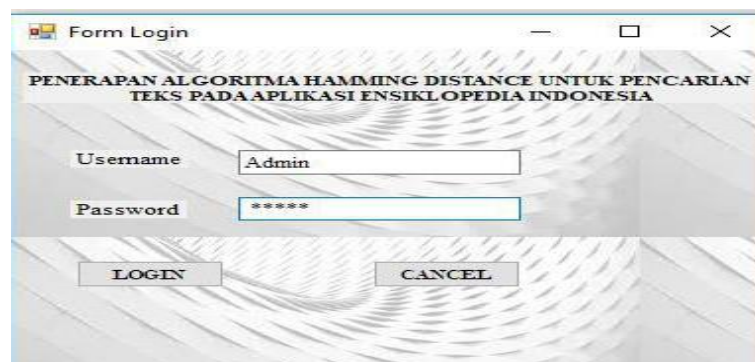
**Tabel 1.** Jarak Hamming

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

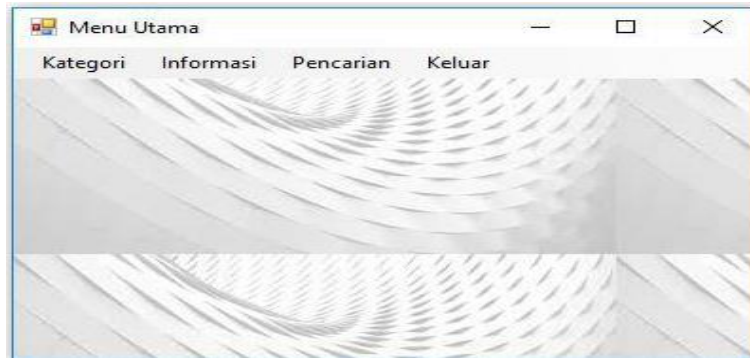
Dari contoh di atas, terlihat bahwa  $C_n n = C_{4,4} = 2$  artinya jarak *Hamming* antara teks  $x$  dan  $y$  adalah 2.

### 3.2 Implementasi

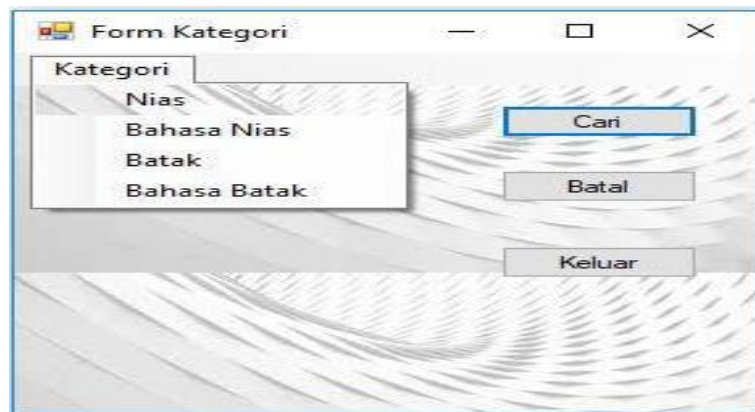
Setelah menyelesaikan semua komponen komputer yang mendukung proses sistem Setelah instalasi, proses selanjutnya adalah implementasi atau penerapan sistem.



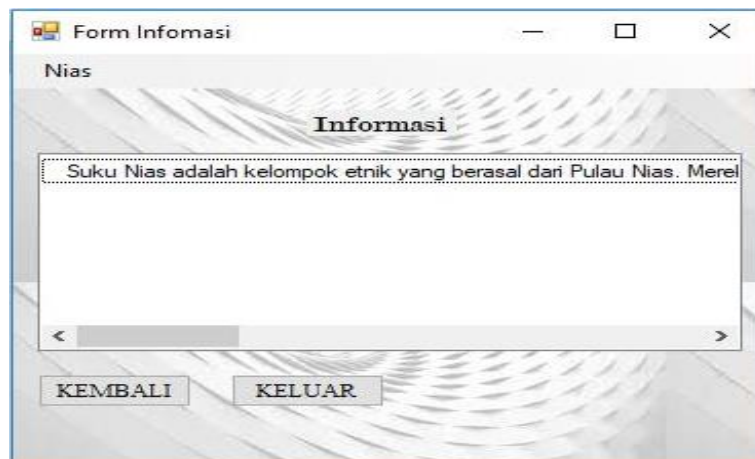
**Gambar 1.** Tampilan Form Login



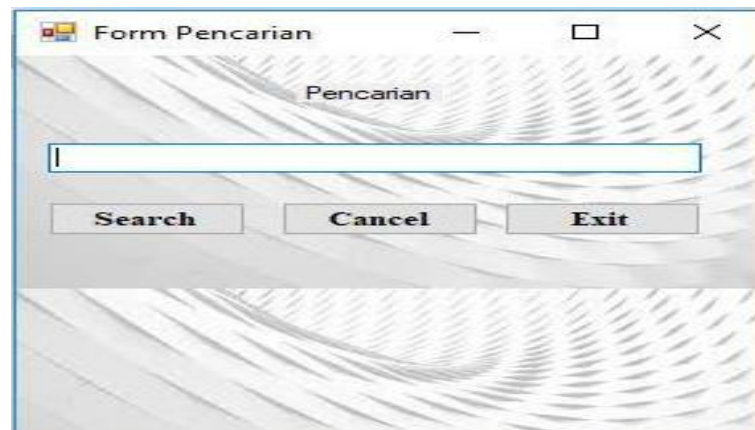
**Gambar 2.** Tampilan Menu Utama



**Gambar 3.** Tampilan Menu Kategori



**Gambar 4.** Tampilan Form Menu Informasi



**Gambar 5.** Tampilan Form Menu Pencarian

#### 4. KESIMPULAN

Berdasarkan analisis dan pembahasan pada bab sebelumnya, penulis dapat menarik kesimpulan yaitu. Dalam pencarian teks pada aplikasi ensiklopedia dapat diterapkan dengan menggunakan algoritma hamming distance. Pencarian teks pada aplikasi ensiklopedia dapat dirancang menggunakan *Visual Studio 2008* dan *mySQL* sebagai *database*.

#### REFERENCES

- [1] M. Dahria, "Kecerdasan Buatan (Artificial Intelligence)," *J. SAINTIKOM*, vol. 5, pp. 185–196.
- [2] E. Buulolo, "IMPLEMENTASI ALGORITMA STRING MATCHING DALAM AM Pencarian SURAT DAN AYAT DALAM BIBLE BERBASIS ANDROID," *Pelita Inform. Budi Darma*, vol. III, no. 2301–9425, p. 23, 2013.
- [3] S. H. Andreas Budiman, Dennis Gunawan, "Implementasi Algoritma Hamming Distance dan Brute Force dalam Mendeteksi Kemiripan Source Code Bahasa Pemrograman C," *ULTIMATICS*, vol. III, no. 2085–4552, p. 72, 2016.
- [4] "No Title," [https://id.wikipedia.org/wiki/Algoritme\\_pencarian](https://id.wikipedia.org/wiki/Algoritme_pencarian).
- [5] E. IMAWATI, "PENGARUH PEMBELAJARAN BERBASIS TEKS TERHADAP KEMAMPUAN MENULIS TEKS DESKRIPTIF," *e - J u r n a l L I T E R A S I*, vol. 1, 2017.
- [6] " <https://id.wikipedia.org/wiki/Aplikasi>.
- [7] " <https://id.wikipedia.org/wiki/Ensiklopedia>.
- [8] S. T. M. K. Sugiarti, Yuni, *Analisis dan Perancangan UML (Unified Modeling Language)*. Yogyakarta: Graha Ilmu, 2013.
- [9] K. Darmayuda, *Pemrograman Aplikasi Database dengan Microsoft Visual Basic.NET 2008*. Bandung: Informatika, 2010.