

Implementasi Algoritma Winnowing Pada Aplikasi Ensiklopedia Alam

Syaiful

Fakultas Ilmu Komputer dan Teknologi Informasi, Prodi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia

Email: syaifulubd20@gmail.com

Abstrak—Ensiklopedia merupakan rangkuman yang berisi informasi terhadap bidang ilmu yang dibahas. Alam merupakan segala sesuatu materi yang terdapat di muka bumi ini, baik itu geologi ataupun makhluk hidup yang terdapat didalamnya. Tetapi di zaman yang modern ini masih belum terdapatnya sebuah aplikasi khusus yang membantu untuk mengakses pengetahuan juga informasi yang berkaitan dengan alam tersebut. Mencari informasi tentang alam melalui browsing harus melalui halaman website yang berbeda – beda hal ini sangat tidak efektif dikarenakan memakan waktu dan tentunya tidaklah efisien. Penyelesaiannya masalah dengan aplikasi pencarian Ensiklopedia Alam yang didalamnya menggunakan algoritma pencocokan string (*string matching*) yaitu algoritma yang dapat mencari kata (*pattern*) dalam sebuah kalimat teks (*string*) untuk menemukan suatu kata dan menampilkan sebuah kata yang ditemukan dalam teks. Algoritma Winnowing digunakan dalam deteksi kesamaan menggunakan fungsi hashing. Winnowing adalah algoritma yang digunakan untuk melakukan proses document fingerprinting.

Kata Kunci: String Matching; Ensiklopedia; Alam; Algoritma Winnowing

Abstract—Encyclopedia is a summary that contains information on the field of science being discussed. Nature is everything that exists on this earth, be it geology or the living things in it. But in this modern era, there is still no special application that helps to access knowledge and information related to nature. Searching for information about nature through browsing has to go through different web pages, this is not very effective because it takes time and is certainly not efficient. The solution to the problem is with the Natural Encyclopedia search application which uses a string matching algorithm, which is an algorithm that can search for words (*patterns*) in a text sentence (*string*) to find a word and display a word found in the text. Winnowing algorithm is used in similarity detection using hashing function. Winnowing is an algorithm used to perform the document fingerprinting process.

Keywords: String Matching; Encyclopedia; Nature; Winnowing Algorithm

1. PENDAHULUAN

Seluruh informasi dan pengetahuan dapat dirangkum pada sebuah referensi yang bernama ensiklopedia. Ensiklopedia merupakan sebuah rangkuman yang berisi informasi tentang cabang ilmu yang dibahas. Ensiklopedia berbeda dengan kamus ataupun yang lainnya dikarenakan pada ensiklopedia menyediakan informasi yang lebih lengkap. Seperti ensiklopedia alam berarti menyediakan segala informasi yang berkaitan tentang alam.

Alam merupakan segala sesuatu materi yang terdapat di muka bumi ini, baik itu geologi ataupun makhluk hidup yang terdapat didalamnya. Banyak jenis – jenis dan kategori tentang alam berdasarkan dengan kelasnya masing – masing, seperti alam semesta yang merupakan segala sesuatu yang berkaitan dengan seluruh dunia, alam sekitar segala sesuatu kondisi dan keadaan juga kehidupan yang berkaitan dengan sekitar kita dan masih banyak lagi yang tersedia di alam seperti kehidupan, energi, sumber daya, flora, fauna dan lain-lain. Banyak ilmu dan pelajaran yang bisa didapatkan dari mempelajari segala sesuatu yang terjadi di alam ini.

Pengetahuan ataupun informasi tentang alam sangat berguna bagi kalangan masyarakat umum. Dengan mengetahui segala informasi – informasi maka akan bertambah wawasan bagi manusia. Khususnya bagi orang tua guna memberikan edukasi terhadap anak – anaknya tentang alam ini. Tetapi di zaman yang modern ini masih belum terdapatnya sebuah aplikasi khusus yang membantu untuk mengakses pengetahuan juga informasi yang berkaitan dengan alam tersebut. Belum adanya aplikasi khusus untuk pencarian informasi tentang alam membuat harus mencari sendiri dengan melakukan pencarian pada media internet atau sering disebut dengan *browsing*. Mencari informasi tentang alam melalui *browsing* harus melalui halaman *website* yang berbeda – beda hal ini sangat tidak efektif dikarenakan memakan waktu dan tentunya tidaklah efisien. Selain itu, mencari informasi tentang alam tersebut dengan *browsing* diharuskan memiliki koneksi internet dan juga biaya. Dalam menyelesaikan permasalahan dibutuhkan berupa aplikasi yang didalamnya menampung seluruh informasi tentang alam dan memudahkan didalam pencariannya. Selain itu, ensiklopedia pada saat ini masih tersedia hanya pada sebuah cetakan buku yang dimana untuk mendapatkannya harus membeli buku ensiklopedia tersebut. Seharusnya pada masa perkembangan teknologi sekarang ini proses pencarian sudah tersedia pada sebuah wadah aplikasi.

Penyelesaiannya masalah dengan aplikasi pencarian Ensiklopedia Alam yang didalamnya menggunakan algoritma *string matching* yang merupakan algoritma untuk mencari kata (*pattern*) pada kalimat teks (*string*) dan menemukan kata serta menampilkan kata yang dalam teks. *String matching* merupakan suatu teknik dalam menentukan keakuratan dari suatu pola yang diberikan[1]. *String Matching* juga berupa teknik menemukan *string* yang cocok dengan pola kira-kira (bukan tepatnya). Dengan kata lain, pencocokan *string* adalah jenis pencarian yang akan menemukan kecocokan bahkan ketika pengguna salah mengeja kata atau memasukkan hanya sebagian kata untuk pencarian. *String Matching* merupakan pencocokan teks yang dicari terhadap *pattern* yang disediakan. Banyak metode ataupun algoritma yang digunakan pada pencocokan *string* saat ini. Salah satu dari algoritma pada *string matching* yang sering digunakan tersebut adalah algoritma *Winnowing*.

Winnowing merupakan algoritma yang dipergunakan pada proses *document fingerprinting* dengan memanfaatkan fungsi *hashing* dari teks. Pada algoritma winnowing dilakukan prosesnya perhitungan nilai hash setiap

k-gram, guna mendapatkan nilai *hash* kemudian menggunakan fungsi rolling hash. Setelah itu di bentuk *window* dari nilai *hash* tersebut. Pada setiap *window* kemudian dipilih nilai *hash* yang terkecil. Jika terdapat lebih dari satu nilai *hash* yang terkecil, maka nilai *hash* yang paling kanan yang digunakan untuk pembentukan *window*. Setelah itu seluruh nilai *hash* yang dipilih kemudian disimpan dijadikan *fingerprint* dari dokumen. *Fingerprint* ini nantinya yang akan dijadikan dasar pembandingan kesamaan antara teks yang telah dimasukan.

Pada penelitian sebelumnya yang dilakukan oleh Agus Setiawan dengan judul “Implementasi Algoritma *Winnowing* Untuk Deteksi Kemiripan Judul Skripsi” mengatakan bahwa algoritma *winnowing* dapat digunakan untuk pencarian kemiripan pada sebuah teks[2].

Pada penelitian lainnya yang telah dilakukan oleh Jarwati dkk dengan judul “Penerapan Algoritma *Winnowing* Pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus Program Studi Sistem Informasi)” mengatakan bahwa algoritma *winnowing* di sistem rekomendasi tersebut mendapatkan hasil yang baik dimana prosesnya membandingkan kata kunci yang ada dan dilihat dari tingkat *persentase* kesamaan kata kunci yang tertinggi[3].

2. METODOLOGI PENELITIAN

2.1 String Matching

Menurut Riyanto Sarno, dkk *String Matching* adalah proses pencarian semua kemunculan query yang selanjutnya disebut *pattern* ke dalam string yang lebih panjang (teks). *Pattern* di lambangkan dengan $x=x[0..m-1]$ dan panjangnya adalah m . Teks dilambangkan dengan $y=y[0..n-1]$ dan panjangnya adalah n . Kedua string terdiri dari sekumpulan karakter yang disebut alfabet yang dilambangkan dengan Σ dan mempunyai ukuran σ . *String matching* dibagi menjadi dua, yakni *exact matching* dan *heuristic* atau *statistical matching*[4].

String matching menurut Ernawati matching adalah sebuah algoritme yang digunakan dalam pencocokan suatu pola kata tertentu terhadap suatu kalimat atau teks panjang. Pencocokan string merupakan bagian penting dari sebuah proses pencarian string (*string searching*) dalam sebuah dokumen. Hasil dari pencarian sebuah string dalam dokumen tergantung dari teknik atau cara pencocokan string yang digunakan[5].

Menurut pendapat Wistiani Astuti, String adalah urutan dari karakter, yang mana karakter ini dapat terdiri dari beberapa alphabet. Misalnya adalah string biner yang terdiri dari dua alphabet, yaitu 0 dan 1, jadi string biner merupakan suatu urutan karakter 0 maupun 1. Contoh yang lain adalah string ASCII yang terdiri dari 256 alphabet. String searching pada dasarnya adalah mencari pola P yang memiliki panjang m dalam suatu teks T yang memiliki panjang n . Baik pola maupun teks dinyatakan dalam array, pola dinyatakan dengan $P[0..m-1]$ dan teks dinyatakan dengan $T[0..n-1]$. Kecocokan adalah apabila karakter pada teks dan karakter pada pola yang dibandingkan adalah sama, sedangkan ketidakcocokan adalah sebaliknya[6].

Dari pendapat – pendapat mengenai string matching diatas maka dapat ditarik sebuah kesimpulan bahwa string matching merupakan pencarian karakter yang memiliki nilai ASCII, digunakan untuk mencocokkan suatu pola kata tertentu (*pattern*) yang terdapat pada kalimat atau teks yang panjang.

2.2 Algoritma Winnowing

Algoritma *winnowing* adalah salah satu algoritma pencocokan *string*. Pada pendeteksiannya, algoritma *winnowing* harus memenuhi kebutuhan mendasar yaitu[2]:

1. *Whitespace insensitivity* yaitu pencarian kalimat mirip seharusnya tidak terpengaruh oleh spasi, jenis huruf (kapital atau normal), tanda baca dan sebagainya.
2. *Noise suppression* yaitu menghindari penemuan kecocokan dengan panjang kata yang terlalu kecil atau kurang relevan seperti “the” dan bukan merupakan kata yang umum digunakan.
3. *Position independence* yaitu penemuan kesamaan harus tidak bergantung pada posisi kata-kata sehingga kata dengan urutan posisi berbeda masih dapat dikenali jika terjadi kesamaan.
 - a. Pembuangan Karakter yang Tidak Relevan, Pada tahap ini proses yang dilakukan yaitu penghapusan tanda baca, spasi dan simbol-simbol seperti @, #, \$, %, *, (,), !, -, _ , ”, +, >, </ dan sebagainya
 - b. Pembentukan Rangkaian *n-gram*, Pembentukan rangkaian *n-gram* pada algoritma *winnowing* dilakukan dengan cara membentuk rangkaian karakter sepanjang n dari hasil pembuangan karakter yang tidak relevan. Nilai n yang baik tidak terlalu kecil dan juga tidak terlalu besar. Rangkaian *n-gram* pertama dimulai dari karakter ke-1 sampai karakter ke- n dan rangkaian kedua dimulai dari karakter ke-2 sampai ke- $n+1$ dan seterusnya sampai terbentuk rangkaian *n-gram* semua karakter.
 - c. Perhitungan Fungsi *Hash* untuk Setiap *n-gram*, Algoritma *winnowing* menggunakan *rolling hash* untuk menghitung nilai Hash masing-masing rangkaian *gram*. Fungsi *hash* dengan *rolling hash* didefinisikan pada persamaan (1).

$$H = c_1 * b_{k-1} + c_2 * b_{k-2} + c_3 * b_{k-3} + \dots + c_k * b_0 \quad (1)$$

- d. Pembentukan *Window* dari Nilai *Hash*, Algoritma *winnowing* tidak menggunakan semua nilai *hash* dari setiap rangkaian *gram* yang dibentuk. Nilai *hash* yang dibentuk pada tahap sebelumnya akan dibagi ke dalam *window* berukuran w . *Window* pertama berisi nilai *hash* pertama sampai nilai *hash* ke- w . *Window* kedua dibentuk dari

nilai *hash* kedua sampai nilai *hash* ke- $w+1$ dan seterusnya sampai terbentuk *window* dari seluruh nilai *hash*.

- e. Pemilihan *Fingerprint* dari Setiap *Window* Setelah terbentuk *window* dari seluruh nilai *hash*, tahap selanjutnya adalah menentukan nilai *fingerprint* teks.

3. HASIL DAN PEMBAHASAN

Ensiklopedia merupakan sebuah rangkuman yang bersisi informasi tentang cabang ilmu yang dibahas. Ensiklopedia bebedra dengan kamus ataupun yang lainnya dikarenakan pada ensiklopedia menyediakan informasi yang lebih lengkap. Seperti ensiklopedia alam berarti menyediakan segala informasi yang berkaitan tentang alam.

Alam merupakan segala sesuatu materi yang terdapat dimuka bumi ini, baik itu geologi ataupun makhluk hidup yang terdapat didalamnya. Banyak jenis – jenis dan kategori tentang alam berdasarkan dengan kelasnya masing – masing. Banyak ilmu dan pelajaran yang bisa didapatkan dari mempelajari segala sesuatu yang terjadi di alam ini.

Pengetahuan ataupun informasi tentang alam sangat berguna bagi kalangan masyarakat umum. Dengan mengetahui segala informasi – informasi maka akan bertambah wawasan bagi manusia. Tetapi di jaman yang modern ini masih belum terdapatnya sebuah aplikasi khusus yang membantu untuk mengakses pengetahuan juga informasi yang berkaitan dengan alam tersebut. Mencari informasi tentang alam melalui *browsing* harus melalui halaman *website* yang berbeda – beda hal ini sangat tidak efektif dikarenakan memakan waktu dan tentunya tidaklah efisien.

Guna menyelesaikan permasalahan tersebut dibutuhkan aplikasi yang didalamnya menampung seluruh informasi tentang alam dan memudahkan didalam pencariannya. Penyelesaiannya masalah dengan aplikasi pencarian Ensiklopedia Alam yang didalamnya menerapkan algoritma *string matching*. Salah satu dari algoritma pada *string matching* yang sering digunakan tersebut adalah algoritma *Winnowing*.

Winnowing merupakan algoritma yang dipergunakan pada proses *document fingerprinting* dengan memanfaatkan fungsi *hashing* dari teks. Pada algoritma *winnowing* dilakukan prosesnya perhitungan nilai *hash* setiap *k*-gram, guna mendapatkan nilai *hash* kemudian menggunakan fungsi rolling hash. Setelah itu di bentuk *window* dari nilai *hash* tersebut. Pada setiap *window* kemudian dipilih nilai *hash* yang terkecil. Jika terdapat lebih dari satu nilai *hash* yang terkecil, maka nilai *hash* yang paling kanan yang digunakan untuk pembentukan *window*. Setelah itu seluruh nilai *hash* yang dipilih kemudian disimpan dijadikan *fingerprint* dari dokumen. *Fingerprint* ini nantinya yang akan dijadikan dasar pembandingan kesamaan antara teks yang telah dimasukan.

3.1 Penerapan Algoritma Winnowing

Dalam melakukan proses pada algoritma *winnowing* dengan mencocokkan nilai *Hash*, oleh sebab itu harus mengetahui terlebih dahulu nilai ASCII yang menjadi *Text* dan juga *Pattern* pada pencocokan string yang akan dicari. Adapun kode *Text* yang akan digunakan adalah “BURUNG ELANG JAWA”, sehingga dapat dilihat kode ASCII pada tabel dibawah ini.

Tabel 1. Kode Kode ASCII

No	Char	Kode ASCII
1	A	065
2	B	066
3	E	069
4	G	071
5	J	074
6	L	076
7	N	078
8	R	082
9	U	085
10	W	087

Pada penelitian ini penulis akan melakukan melakukan pencarian dengan teks AL-BAQARAH yang menggunakan kata “ELANG JAWA” sebagai *pattern*-nya maka proses akan terbentuk sebagai berikut.

Text : BURUNG ELANG JAWA

Pattern : ELANG JAWA

1. Langkah Pertama :

Text yang digunakan pada penelitian ini adalah “BURUNG ELANG JAWA” dan *pattern* adalah “ELANG JAWA” pada langkah pertama menghapuskan spasi atau tanda baca yang tidak relevan.

Text : BURUNGELANGJAWA

Pattern : ELANGJAWA

2. Langkah Kedua:

Selanjutnya adalah pembagian N-gram, disini kata dipecah menjadi potongan kata dan pada setiap potongan kata terdapat karakter sebanyak N. Berikut ini adalah contoh proses parsing N-garm N=4: Untuk pertama, membuat N-garm N=4 untuk teks seperti dibawah ini

```

B  U  R  U
   U  R  U  N
      R  U  N  G
         U  N  G  E
            N  G  E  L
               G  E  L  A
                  E  L  A  N
                     L  A  N  G
                        A  N  G  J
                           N  G  J  A
                              G  J  A  W
                                 J  A  W  A

```

Setelah membuat N-gram untuk teks, selanjutnya membuat N-garm N=4 untuk pattern seperti dibawah ini.

```

E  L  A  N
   L  A  N  G
      A  N  G  J
         N  G  J  A
            G  J  A  W
               J  A  W  A

```

3. Langkah ketiga :

Cara menghitung nilai *hash* pada *n-gram* dengan nilai *basis* (b) = 2, panjang rangkaian n-gram(n) = 4 sehingga dapat dihitung untuk nilai Hash setiap N-gram dan juga *Pattern* dengan menggunakan rumus

A. Nilai Hash N-garm Teks

Untuk nilai Hash N-Garm teks dapat dilihat pada dibawah ini :

1. N-Gram 1 [BURU]

$$\begin{aligned}
 H &= 66 * 2^{4-1} + 85 * 2^{4-2} + 82 * 2^{4-3} + 85 * 2^{4-4} \\
 &= 66 * 2^3 + 85 * 2^2 + 82 * 2^1 + 85 * 2^0 \\
 &= 66 * 8 + 85 * 4 + 82 * 2 + 85 * 1 \\
 &= 528 + 340 + 164 + 85 \\
 &= 1117
 \end{aligned}$$
2. N-Gram 2 [URUN]

$$\begin{aligned}
 H &= 85 * 2^{4-1} + 82 * 2^{4-2} + 85 * 2^{4-3} + 78 * 2^{4-4} \\
 &= 85 * 2^3 + 82 * 2^2 + 85 * 2^1 + 78 * 2^0 \\
 &= 85 * 8 + 82 * 4 + 85 * 2 + 78 * 1 \\
 &= 680 + 328 + 170 + 78 \\
 &= 1256
 \end{aligned}$$
3. N-Gram 3 [RUNG]

$$\begin{aligned}
 H &= 82 * 2^{4-1} + 85 * 2^{4-2} + 78 * 2^{4-3} + 71 * 2^{4-4} \\
 &= 82 * 2^3 + 85 * 2^2 + 78 * 2^1 + 71 * 2^0 \\
 &= 82 * 8 + 85 * 4 + 78 * 2 + 71 * 1 \\
 &= 656 + 340 + 156 + 71 \\
 &= 1223
 \end{aligned}$$
4. N-Gram 4 [UNGE]

$$\begin{aligned}
 H &= 85 * 2^{4-1} + 78 * 2^{4-2} + 71 * 2^{4-3} + 69 * 2^{4-4} \\
 &= 85 * 2^3 + 78 * 2^2 + 71 * 2^1 + 69 * 2^0 \\
 &= 85 * 8 + 78 * 4 + 71 * 2 + 69 * 1 \\
 &= 680 + 312 + 142 + 69 \\
 &= 1203
 \end{aligned}$$
5. N-Gram 5 [NGEL]

$$\begin{aligned}
 H &= 78 * 2^{4-1} + 71 * 2^{4-2} + 69 * 2^{4-3} + 76 * 2^{4-4} \\
 &= 78 * 2^3 + 71 * 2^2 + 69 * 2^1 + 76 * 2^0 \\
 &= 78 * 8 + 71 * 4 + 69 * 2 + 76 * 1 \\
 &= 624 + 284 + 138 + 76 \\
 &= 1122
 \end{aligned}$$
6. N-Gram 6 [GELA]

$$\begin{aligned}
 H &= 71 * 2^{4-1} + 69 * 2^{4-2} + 76 * 2^{4-3} + 65 * 2^{4-4} \\
 &= 71 * 2^3 + 69 * 2^2 + 76 * 2^1 + 65 * 2^0
 \end{aligned}$$

$$\begin{aligned}
&= 71 * 8 + 69 * 4 + 76 * 2 + 65 * 1 \\
&= 568 + 276 + 152 + 65 \\
&= 1061
\end{aligned}$$

7. N-Gram 7 [ELAN]

$$\begin{aligned}
H &= 69 * 2^{4-1} + 76 * 2^{4-2} + 65 * 2^{4-3} + 78 * 2^{4-4} \\
&= 69 * 2^3 + 76 * 2^2 + 65 * 2^1 + 78 * 2^0 \\
&= 69 * 8 + 76 * 4 + 65 * 2 + 78 * 1 \\
&= 552 + 304 + 130 + 78 \\
&= 1064
\end{aligned}$$

8. N-Gram 8 [LANG]

$$\begin{aligned}
H &= 76 * 2^{4-1} + 65 * 2^{4-2} + 78 * 2^{4-3} + 71 * 2^{4-4} \\
&= 76 * 2^3 + 65 * 2^2 + 78 * 2^1 + 71 * 2^0 \\
&= 76 * 8 + 65 * 4 + 78 * 2 + 71 * 1 \\
&= 608 + 260 + 156 + 71 \\
&= 1095
\end{aligned}$$

9. N-Gram 9 [ANGJ]

$$\begin{aligned}
H &= 65 * 2^{4-1} + 78 * 2^{4-2} + 71 * 2^{4-3} + 74 * 2^{4-4} \\
&= 65 * 2^3 + 78 * 2^2 + 71 * 2^1 + 74 * 2^0 \\
&= 65 * 8 + 78 * 4 + 71 * 2 + 74 * 1 \\
&= 520 + 312 + 142 + 74 \\
&= 1048
\end{aligned}$$

10. N-Gram 10 [NGJA]

$$\begin{aligned}
H &= 78 * 2^{4-1} + 71 * 2^{4-2} + 74 * 2^{4-3} + 65 * 2^{4-4} \\
&= 78 * 2^3 + 71 * 2^2 + 74 * 2^1 + 65 * 2^0 \\
&= 78 * 8 + 71 * 4 + 74 * 2 + 65 * 1 \\
&= 624 + 284 + 148 + 65 \\
&= 1121
\end{aligned}$$

11. N-Gram 11 [GJAW]

$$\begin{aligned}
H &= 71 * 2^{4-1} + 74 * 2^{4-2} + 65 * 2^{4-3} + 87 * 2^{4-4} \\
&= 71 * 2^3 + 74 * 2^2 + 65 * 2^1 + 87 * 2^0 \\
&= 71 * 8 + 74 * 4 + 65 * 2 + 87 * 1 \\
&= 568 + 296 + 130 + 87 \\
&= 1081
\end{aligned}$$

12. N-Gram 12 [JAWA]

$$\begin{aligned}
H &= 74 * 2^{4-1} + 65 * 2^{4-2} + 87 * 2^{4-3} + 65 * 2^{4-4} \\
&= 74 * 2^3 + 65 * 2^2 + 87 * 2^1 + 65 * 2^0 \\
&= 74 * 8 + 65 * 4 + 87 * 2 + 65 * 1 \\
&= 592 + 260 + 174 + 65 \\
&= 1091
\end{aligned}$$

B. Nilai Hash N-garm Pattern

1. N-Gram 1 [ELAN]

$$\begin{aligned}
H &= 69 * 2^{4-1} + 76 * 2^{4-2} + 65 * 2^{4-3} + 78 * 2^{4-4} \\
&= 69 * 2^3 + 76 * 2^2 + 65 * 2^1 + 78 * 2^0 \\
&= 69 * 8 + 76 * 4 + 65 * 2 + 78 * 1 \\
&= 552 + 304 + 130 + 78 \\
&= 1064
\end{aligned}$$

2. N-Gram 2 [LANG]

$$\begin{aligned}
H &= 76 * 2^{4-1} + 65 * 2^{4-2} + 78 * 2^{4-3} + 71 * 2^{4-4} \\
&= 76 * 2^3 + 65 * 2^2 + 78 * 2^1 + 71 * 2^0 \\
&= 76 * 8 + 65 * 4 + 78 * 2 + 71 * 1 \\
&= 608 + 260 + 156 + 71 \\
&= 1095
\end{aligned}$$

3. N-Gram 3 [ANGJ]

$$\begin{aligned}
H &= 65 * 2^{4-1} + 78 * 2^{4-2} + 71 * 2^{4-3} + 74 * 2^{4-4} \\
&= 65 * 2^3 + 78 * 2^2 + 71 * 2^1 + 74 * 2^0 \\
&= 65 * 8 + 78 * 4 + 71 * 2 + 74 * 1 \\
&= 520 + 312 + 142 + 74 \\
&= 1048
\end{aligned}$$

4. N-Gram 4 [NGJA]

$$\begin{aligned}
H &= 78 * 2^{4-1} + 71 * 2^{4-2} + 74 * 2^{4-3} + 65 * 2^{4-4} \\
&= 78 * 2^3 + 71 * 2^2 + 74 * 2^1 + 65 * 2^0 \\
&= 78 * 8 + 71 * 4 + 74 * 2 + 65 * 1
\end{aligned}$$

$$= 624 + 284 + 148 + 65$$

$$= 1121$$

5. N-Gram 5 [GJAW]

$$H = 71 * 2^{4-1} + 74 * 2^{4-2} + 65 * 2^{4-3} + 87 * 2^{4-4}$$

$$= 71 * 2^3 + 74 * 2^2 + 65 * 2^1 + 87 * 2^0$$

$$= 71 * 8 + 74 * 4 + 65 * 2 + 87 * 1$$

$$= 568 + 296 + 130 + 87$$

$$= 1081$$

6. N-Gram 6 [JAWA]

$$H = 74 * 2^{4-1} + 65 * 2^{4-2} + 87 * 2^{4-3} + 65 * 2^{4-4}$$

$$= 74 * 2^3 + 65 * 2^2 + 87 * 2^1 + 65 * 2^0$$

$$= 74 * 8 + 65 * 4 + 87 * 2 + 65 * 1$$

$$= 592 + 260 + 174 + 65$$

$$= 1091$$

Setelah dilakukan perhitungan nilai Hash, maka didapatkan nilai Hash N-garm *Teks* 1 sampai dengan K-garm *Teks* 12 dengan nilai Hash [1117, 1256, 1223, 1203, 1122, 1061, 1064, 1095, 1048, 1121, 1081, 1091]

Setelah dilakukan perhitungan nilai Hash, maka didapatkan nilai Hash N-garm *Pattern* 1 sampai dengan N-garm *Pattern* 6 dengan nilai Hash [1064, 1095, 1048, 1121, 1081, 1091]

4. Langka Keempat

Window dari perhitungan nilai *hash text* ditahap sebelumnya dengan lebar *window* (w) = 3, *window* BURUNG ELANG JAWA yaitu:

[1117, 1256, 1223]
 [1256, 1223, 1203]
 [1223, 1203, 1122]
 [1203, 1122, 1061]
 [1122, 1061, 1064]
 [1061, 1064, 1095]
 [1064, 1095, 1048]
 [1095, 1048, 1121]
 [1048, 1121, 1081]
 [1121, 1081, 1091]

Window dari perhitungan nilai *hash pattern* dari tahap sebelumnya dengan lebar *window* (w) = 3, *window* ELANG JAWA yaitu:

[1064, 1095, 1048]
 [1095, 1048, 1121]
 [1048, 1121, 1081]
 [1121, 1081, 1091]

5. Langka Kelima

Pemilihan nilai *fingerprint* dari hasil pembentukan *window* pada tahap sebelumnya, *fingerprint teks* yaitu:

[1117, 1256, 1223] = 1117
 [1256, 1223, 1203] = 1203
 [1223, 1203, 1122] = 1203
 [1203, 1122, 1061] = 1203
 [1122, 1061, 1064] = 1061
 [1061, 1064, 1095] = 1061
 [1064, 1095, 1048] = 1048
 [1095, 1048, 1121] = 1048
 [1048, 1121, 1081] = 1048
 [1121, 1081, 1091] = 1081

Pemilihan nilai *fingerprint* dari hasil pembentukan *window* pada tahap sebelumnya, *fingerprint pattern* yaitu:

[1064, 1095, 1048] = 1048
 [1095, 1048, 1121] = 1048
 [1048, 1121, 1081] = 1048
 [1121, 1081, 1091] = 1081

Fingerprint yang terbentuk yaitu :

Fingerprint Teks

1117, 1203, 1061, 1048, 1081

Fingerprint Pattern

1048, 1081

6. Langkah Keenam

Perhitungan kesamaan dengan menggunakan persamaan jaccard coefficient yaitu:

Fingerprint Teks

1117, 1203, 1061, 1048, 1081

Fingerprint Pattern

1048, 1081

Jumlah fingerprint yang sama (1048, 1081) = 2

Keseluruhan *fingerprint* =(1117, 1203, 1061, 1048, 1081) = 5

Proses selanjutnya menghitung *similitary* yaitu tingkat kesamaan antara *text* dan *pattern*, yaitu berapa persen tingkat kesamaannya. Berikut ini adalah proses menghitung *similitary* dua dokumen diatas :

$$\begin{aligned}\text{Similitary (Text,Pattern)} &= \frac{\sum H_{\text{text}} \cap \sum H_{\text{pattern}}}{\sum H_{\text{text}} \cup \sum H_{\text{pattern}}} \times 100\% \\ &= \frac{2}{2+5-2} * 100\% \\ &= \frac{2}{5} * 100\% \\ &= 0,40 * 100\% = 40\%\end{aligned}$$

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan tentang aplikasi ensiklopedia alam, maka dapat diambil kesimpulan berupa hasil akhir dari penulisan yaitu Proses pencarian ensiklopedia alam dapat dilakukan dengan cara melakukan inputan kata kunci yang ingin dicari. Lalu aplikasi akan menampilkan informasi – informasi berdasarkan dengan kata kunci yang dicari. Dengan penerapan algoritma winnowing dapat mempercepat proses pencarian informasi berdasarkan tingkat kemiripan terhadap kata kunci

REFERENCES

- [1] A. A. B. Ginting and D. P. Utomo, "Perancangan Aplikasi Catalog Wisata Di Sumatera Utara Menggunakan Algoritma Rabin-Karp," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 57–63, 2019, doi: 10.30865/komik.v3i1.1568.
- [2] A. R. Syahputra, "Implementasi Algoritma Winnowing Untuk Deteksi Kemiripan Judul Skripsi Studi Kasus: STMIK Budi Darma," *Pelita Inform. Budi Darma*, vol. 12, no. 1, pp. 1–9, 2017.
- [3] J. Jarwati, Antonius Cahya Prihandoko, and W. E. Y. R., "Penerapan Algoritma Winnowing Pada Sistem Rekomendasi Penentuan Dosen Pembimbing Skripsi (Studi Kasus Program Studi Sistem Informasi)," *Berk. SAINSTEK*, vol. V, no. 1, pp. 11–20, 2017.
- [4] Riyanto Sarno; Yenni Anistyasari; Fitri Rahimi, *Semantic Search*. Yogyakarta: Andi, 2012.
- [5] E. Ernawati, A. Johar, and S. Setiawan, "Implementasi Metode String Matching Untuk Pencarian Berita Utama Pada Portal Berita Berbasis Android (Studi Kasus: Harian Rakyat Bengkulu)," *Pseudocode*, vol. 6, no. 1, pp. 77–82, 2019, doi: 10.33369/pseudocode.6.1.77-82.
- [6] W. Astuti, "Analisis String Matching Pada Judul Skripsi Dengan Algoritma Knuth-Morris Pratt (Kmp)," *Ilk. J. Ilm.*, vol. 9, no. 2, pp. 167–172, 2017, doi: 10.33096/ilkom.v9i2.136.167-172.
- [7] W. E. Bebas, "Ensiklopedia," *Wikipedia Ensiklopedia Bebas*, 2020. <https://id.wikipedia.org/wiki/Ensiklopedia>.
- [8] W. E. Bebas, "Alam," *Wikipedia Ensiklopedia Bebas*, 2020. <https://id.wikipedia.org/wiki/Alam>.
- [9] P. Sulistyorini, "Pemodelan Visual dengan Menggunakan UML dan Rational Rose," *J. Teknol. Inf. Din. Vol.*, vol. XIV, no. 1, pp. 23–29, 2009.
- [10] R. A. S. M. Shalahuddin, *Rekayasa Perangkat Lunak (Arsitektur Dan Berorientasi Objek)*. Bandung: Informatika, 2014.
- [11] Hendrayudi, *VB 2008 untuk Berbagai Keperluan Pemrograman*. Jakarta: PT. Elex Media Komputindo, 2009.
- [12] P. A. dan S. (PAS), *Membuat Aplikasi Client Server dengan Visual Basic 2008*. Yogyakarta: Andi, 2010.