

Aplikasi *String Matching* dalam Pengkategorian Penyebab Mutasi Kromosom

Mochamad Try Yulyanto (13513067)

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia

mochamadtry@students.itb.ac.id

Abstract— *Mutasi berasal dari kata mutates (bahasa latin) yang artinya adalah perubahan. Mutasi didefinisikan sebagai perubahan materi genetik yang dapat diwariskan secara genetik pada keturunannya. Agen yang menyebabkan mutasi disebut mutagen. Makhluk hidup yang mengalami mutasi disebut mutan. Perubahan susunan genetik ini menyebabkan terjadinya perubahan gen dan akhirnya menyebabkan perubahan alel dan fenotip makhluk hidup. Salah satu tingkat mutasi adalah mutasi kromosom. Kromosom sendiri tersusun atas rantai DNA. Dengan DNA disimbolkan dengan kode genetik yang terdiri atas string. Penyimpangan yang terjadi dalam kromosom dapat dideteksi menggunakan pencocokan string terhadap data kromosom yang tidak menyimpang. Dari hasil pencocokan string tersebut maka mutasi pada kromosom dapat diketahui*

Index Terms—Kromosom, Delesi, Duplikasi, inversi, Translokasi, Mutasi.

I. PENDAHULUAN

Dalam kehidupan sehari-hari kita sering mendengar istilah rekayasa genetika, dan penyakit-penyakit yang berhubungan dengan kerusakan atau kelainan kromosom. Kita pun sering melihat atau memakan buah jambu batu berukuran besar atau buah semangka tanpa biji. Materi genetik buah-buahan tersebut secara sengaja telah diubah oleh manusia. Perubahan materi genetik juga bisa terjadi secara alami baik disebabkan oleh gangguan fisik atau kimia yang menyebabkan kesalahan dalam pembelahan sel. Perubahan materi genetik dapat terjadi pada tingkat DNA maupun kromosom.

Mutasi kromosom merupakan mutasi atau perubahan materi genetik yang disebabkan oleh perubahan pada kromosom. Kromosom sendiri merupakan suatu struktur padat yang terdiri dari dua komponen molekul, yaitu protein dan DNA. Struktur padat kromosom hanya dapat terlihat dengan jelas pada saat pembelahan sel. Perubahan jumlah dan susunan kromosom berkaitan dengan sejumlah kelainan serius pada manusia. Perubahan susunan kromosom dapat disebabkan oleh delesi (penghilangan),

duplikasi (penggandaan), inversi (pembalikan) dan translokasi.

Seperti yang telah dibahas sebelumnya, kromosom tersusun atas DNA dan protein. DNA merupakan molekul panjang yang menyimpan informasi genetik. Total informasi genetik yang disimpan dalam DNA suatu sel disebut genom. Rangkaian gugus molekul dari DNA digambarkan dengan untaian string basa nitrogen yang terdiri atas basa purin (adenine = A dan guanine = G) dan basa pirimidin (sitosin = C dan Timin = T). Keempat basa nitrogen tersebut membentuk suatu kode genetik yang dapat dikenali. Jika terdapat ketidakcocokan urutan ataupun kelebihan/kekurangan jumlah dalam untaian DNA tersebut dapat dikatakan telah terjadi perubahan materi genetik. Kejadian tersebut dapat dikenali melalui pencocokan string (*string matching*) dengan 4 kemungkinan mutasi kromosom.

Pada makalah ini akan ditunjukkan bagaimana peranan algoritma *string matching* dalam menentukan kategori mutasi yang terjadi pada kromosom dengan pengenalan terhadap DNA dengan menggunakan algoritma Knuth Morish Pratt dan Boyer-Moore.

II. LANDASAN TEORI

2.1 Pencocokan String (*String Matching*)

Pencocokan string (*string matching*) merupakan permasalahan paling sederhana dari semua permasalahan string lainnya, dan dianggap sebagai bagian dari pemrosesan data, pengkompresian data, analisis leksikal, dan temu balik informasi. Teknik untuk menyelesaikan permasalahan pencocokan string biasanya akan menghasilkan implikasi langsung ke aplikasi string lainnya. Persoalan pencarian string dirumuskan sebagai berikut:

1. teks (*text*), yaitu (*long*) string yang panjangnya n karakter
2. *pattern*, yaitu string dengan panjang m karakter ($m < n$) yang akan dicari di dalam teks.

Carilah (*find* atau *locate*) lokasi pertama di dalam teks yang bersesuaian dengan *pattern*. Aplikasi dari masalah pencocokan string antara lain pencarian suatu kata di dalam dokumen (misalnya menu *Find* di dalam *Microsoft Word*). Algoritma pencocokan string bertujuan untuk

menentukan letak sebuah pattern di dalam text. Ada banyak algoritma yang telah dikembangkan untuk mencari algoritma apa yang terbaik dalam menyelesaikan permasalahan string matching ini, di antaranya Algoritma *Brute Force*, Algoritma KMP dan Algoritma Boyer-Moore.

2.2 Algoritma Knuth-Morris-Pratt

Pada algoritma *brute force*, setiap kali ditemukan ketidakcocokan *pattern* dengan teks, maka *pattern* digeser satu karakter ke kanan. Sedangkan pada algoritma KMP, kita memelihara informasi yang digunakan untuk melakukan jumlah pergeseran. Algoritma menggunakan informasi tersebut untuk membuat pergeseran yang lebih jauh, tidak hanya satu karakter seperti pada algoritma *brute force*.

Dengan algoritma KMP ini, waktu pencarian dapat dikurangi secara signifikan. Misalkan A adalah alfabet dan $x = x_1x_2\dots x_k$, $k \in \mathbf{N}$, adalah *string* yang panjangnya k yang dibentuk dari karakter-karakter di dalam alfabet A . Awalan (*prefix*) dari x adalah upa-string (*substring*) u dengan

$$u = x_1x_2\dots x_{k-1}, k \in \{1, 2, \dots, k-1\}$$

dengan kata lain, x diawali dengan u .

Akhiran (*suffix*) dari x adalah upa-string (*substring*) u . Dengan :

$$u = x_{k-b}x_{k-b+1}\dots x_k, k \in \{1, 2, \dots, k-1\}$$

dengan kata lain, x diakhiri dengan v .

Pinggiran (*border*) dari x adalah upa-string r sedemikian sehingga

$$r = x_1x_2\dots x_{k-1} \text{ dan } u = x_{k-b}x_{k-b+1}\dots x_k, k \in \{1, \dots, k-1\}$$

2.2.1 Fungsi Pinggiran (Border Function)

Fungsi *pinggiran* $b(j)$ didefinisikan sebagai ukuran awalan terpanjang dari P yang merupakan akhiran dari $P[1..j]$. Sebagai contoh, tinjau *pattern* $P = ababaa$. Nilai F untuk setiap karakter di dalam P adalah sebagai berikut:

Table 1 Contoh Fungsi Pinggiran

j	1	2	3	4	5	6
$P[j]$	a	b	a	b	a	a
$b(j)$	0	0	1	2	3	1

2.2.2 Pseudo Code Algoritma KMP

```

procedure KMPsearch(input m, n :
integer, input P : array[1..m] of
char, input T :
array[1..n] of char,
output idx :
integer)
Deklarasi
i, j : integer
ketemu : boolean
b : array[1..m] of integer

```

```

procedure HitungPinggiran(input m :
integer, P : array[1..m] of char,
output b : array[1..m] of integer)
{ Menghitung nilai b[1..m] untuk
pattern P[1..m] }
Algoritma:

HitungPinggiran(m, P, b)
j←0
i←1
ketemu←false
while (i ≤ n and not ketemu) do

    while((j > 0) and (P[j+1]≠T[i])) do
        j←b[j]
    endwhile

    if P[j+1]=T[i] then
        j←j+1
    endif
    if j = m then
        ketemu←true
    else
        i←i+1
    endif
endwhile
if ketemu then
    idx←i-m+1 { catatan: jika indeks array
dimulai dari 0, maka idx←i-m }
else
    idx←-1
endif

```

Table 2 Pseudo Code KMP

2.3 Algoritma Boyer-Moore

Algoritma pencocokan string Boyer Moore didasarkan pada 2 teknik, yaitu :

a) Looking-glass technique

Teknik ini merupakan cara untuk menemukan sebuah pattern dalam teks dengan memulai pencocokan dari akhir string pattern.

b) Character-jumpechnique

Saat terjadi ketidakcocokan, pencarian akan dilanjutkan setelah menggeser pattern sebesar nilai tertentu untuk menghindari pencocokan yang tidak berguna.

2.3.1 Langkah Algoritma Boyer-Moore

Langkah-langkah algoritma Boyer-Moore :

1. Buat tabel pergeseran string yang dicari (S) dengan pendekatan Match Heuristic (MH) dan Occurrence Heuristic (OH), untuk menentukan jumlah pergeseran yang akan dilakukan jika

mendapat karakter tidak cocok pada proses pencocokan dengan string (T).

2. Jika dalam proses perbandingan terjadi ketidakcocokan antara pasangan karakter pada S dan karakter pada T, pergeseran dilakukan dengan memilih salah satu nilai pergeseran dari dua tabel analisa string, yang memiliki nilai pergeseran paling besar.
3. Dua kemungkinan penyelesaian dalam melakukan pergeseran S, jika sebelumnya belum ada karakter yang cocok adalah dengan melihat nilai pergeseran hanya pada tabel occurrence heuristic : Jika karakter yang tidak cocok tidak ada pada S maka pergeseran adalah sebanyak jumlah karakter pada S. dan jika karakter yang tidak cocok ada pada S, maka banyaknya pergeseran bergantung dari nilai pada tabel.
4. Jika karakter pada teks yang sedang dibandingkan cocok dengan karakter pada S, maka posisi karakter pada S dan T diturunkan sebanyak 1 posisi, kemudian lanjutkan dengan pencocokan pada posisi tersebut dan seterusnya. Jika kemudian terjadi ketidakcocokan karakter S dan T, maka pilih nilai pergeseran terbesar dari dua tabel analisa pattern yaitu nilai dari tabel match heuristic dan nilai tabel occurrence heuristic dikurangi dengan jumlah karakter yang telah cocok.
5. Jika semua karakter telah cocok, artinya S telah ditemukan di dalam T, selanjutnya geser pattern sebesar 1 karakter.
6. Lanjutkan sampai akhir string T.

2.3.2 Pseudo Code Algoritma Boyer-Moore

```
BOYER_MOORE_MATCHER
(T, P)
Input:
Text with n characters
and
Pattern with m characters
```

```
Output:
Index of the first
substring of T matching P
Compute function last
i ← m - 1
j ← m - 1
Repeat
    if P[j] = T[i] then
        if j=0 then
            return i
```

```
// we have a match
else
    i ← i - 1
    j ← j - 1
else
    i ← i + m - Min(j, 1 +
        last[T[i]])
    j ← m - 1
until i > n -
Return "no match"
```

Table 3 Pseudo Code Boyer-Moore

2.4 Definisi Kromosom

Kromosom adalah struktur dalam inti sel yang terdiri dari DNA yang terikat dengan histon dan protein lain. Gen terbuat dari DNA (meskipun sebagian dari urutan DNA bukan bagian dari gen apapun). Gen disusun sepanjang kromosom dalam urutan yang kontinu.

Struktur protein kromosom memungkinkan untuk aktivasi selektif (gen ditranskripsi menjadi protein) atau inaktivasi (gen tidak diungkapkan), dan dengan demikian untuk ekspresi diferensial dari genom dalam jenis sel yang berbeda dan ekspresi gen dalam urutan yang tepat selama perkembangan organisme atau pada berbagai kondisi metabolik.

2.5 Tempat Terjadinya Mutasi

Makhluk hidup, individu, atau sel yang mengalami mutasi disebut makhluk, individu, atau sel mutan. Berdasarkan tempat terjadinya, yaitu sel-sel yang mengalami mutasi, mutasi dibedakan menjadi mutasi gametik dan somatik.

a. Mutasi Gametik

Mutasi yang terjadi pada sel gamet. Karena sel gamet mengadakan pembuahan, gamet mutan yang mengalami pembuahan dapat mewariskan mutasi tersebut kepada turunannya. Mutasi gametik bersifat diwariskan dan merupakan mutasi yang dikekalkan secara genetic. Gen-gen yang mengalami mutasi di dalam gamet dapat berupa mutasi autosomal yang bisa menghasilkan mutasi dominan atau mutasi resesif, tergantung pada sifat yang dihasilkannya.

b. Mutasi Somatik

Merupakan mutasi genetik yang terjadi pada sel somatik setelah pembuahan. Mutasi ini dapat menyebabkan berbagai masalah medis, dan orang awam sering mengasosiasikan mereka khusus dengan kanker. Mutasi somatik dapat diidentifikasi dengan memeriksa materi genetik dalam sel yang diragukan dan membandingkannya dengan sel dari tempat lain di tubuh, DNA dalam dua sel akan berbeda, meskipun fakta bahwa itu tidak seharusnya.

Ada dua jenis sel: sel somatik dan sel-sel germinal. Sel germinal akhirnya menimbulkan gamet, sedangkan sel somatik menimbulkan segala sesuatu yang lain. Jika mutasi somatik terjadi di dalam rahim,

semua sel berasal dari sel somatik akan membawa mutasi. Ini hasil dalam situasi yang disebut mosaik genetik, di mana beberapa sel dalam tubuh seseorang memiliki DNA yang berbeda dari sel-sel lain.

2.6 Mekanisme Mutasi

Secara ringkas telah disebutkan bahwa mutasi dapat mempengaruhi DNA maupun kromosom. DNA dapat dipengaruhi pada saat sintesis DNA (replikasi). Pada saat tersebut factor mutagenic mempengaruhi pemasangan basa nukleotida sehingga tidak berpasangan dengan basa neuklotida yang seharusnya (*mismatch*). Misalnya triplet DNA cetakan adalah TTA. Namun, karena adanya mutagen menyebabkan DNA polymerase memasangkan A dengan C bukan dengan T.

Kromosom juga dapat dipengaruhi oleh mutagen pada saat pemakatan DNA dalam kromosom (profase), pemisahan kromatid (contoh pada peristiwa gagal berpisah kromosom saat anaphase), penarikan kromosom oleh benang spindle, dan sintesis dinding sel (sitokinesis) setelah anafase.

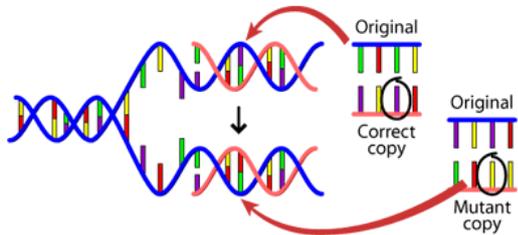


Figure 1 Mekanisme Mutasi

Sumber : <https://i0.wp.com/evolution.berkeley.edu>

2.7 Mutasi Kromosom

Mutasi kromosom adalah perubahan jumlah kromosom dan susunan gen dalam kromosom. Mutasi ini sering terjadi karena adanya kesalahan saat meiosis. Mutasi kromosom dibedakan menjadi dua sebagai berikut.

Mutasi kromosom dapat mengakibatkan perubahan dalam jumlah kromosom dalam sel atau perubahan struktur kromosom. Tidak seperti mutasi gen yang mengubah gen tunggal atau segmen yang lebih besar dari DNA pada kromosom, mutasi kromosom berubah dan berdampak pada seluruh kromosom.

Kromosom, yang mengandung informasi genetik dalam sel-sel dari organisme yang berbeda, dapat menjadi bermutasi melalui berbagai proses yang berbeda, sering merugikan dari organisme bermutasi.

2.8 Kategori Mutasi Kromosom

Perubahan susunan kromosom dapat disebabkan oleh

a. Delesi

Terjadi ketika kromosom kehilangan sebagian segmennya. Defisiensi ini mempunyai pengaruh genetik, antara lain efek letal (kematian) dan pseudodominan (pemunculan fenotipe sifat resesif, seperti sifat dominan).



Figure 2 Contoh Delesi

Sumber : <http://softilmu.blogspot.com>

b. Duplikasi

Duplikasi terjadi jika kromosom memperoleh tambahan sebagian segmen kromosom lainnya. Duplikasi mempunyai efek genetik, antara lain melindungi pengaruh gen resesif yang merugikan untuk evaluasi materi genetik, dan menghasilkan efek posisi (menghasilkan fenotipe baru).



Figure 3 Contoh Duplikasi

Sumber : <http://softilmu.blogspot.com>

c. Inversi

pembalikan urutan-urutan pada susunan gen. Inversi berperan menekan terjadinya peristiwa pindah silang.



Figure 4 Contoh Inversi

Sumber : <http://softilmu.blogspot.com>

d. Translokasi

Translokasi adalah pertukaran sebagian kromosom dengan kromosom nonhomolog lainnya sehingga menghasilkan efek posisi.

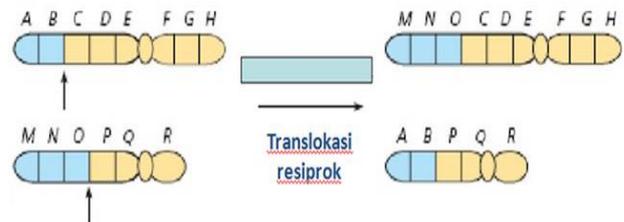


Figure 5 Contoh Translokasi

Sumber : <http://softilmu.blogspot.com>

III. HASIL PENGKATEGORIAN MUTASI

Pada testing pengkategorian mutasi hanya akan dites dalam kategori Tidak terjadi mutasi, Delesi, Duplikasi dan Inversi saja. Karena ketiga kategori itu yang cocok di test menggunakan *string matching*.

- Tidak terjadi mutasi kromosom – KMP

```

Command Prompt
F:\Kuliah\StiMa\Tugas Besar III\Setor>java test
Kromoson : ABCDEFG
Cek Kromoson : ABCDEFG
Tidak Terjadi Mutasi Kromosom
Jumlah Langkah : 8
  
```

Table 4 Tidak Mutasi - KMP

- Tidak terjadi mutasi kromosom – Boyer Moore

```

Command Prompt
F:\Kuliah\StiMa\Tugas Besar III\Setor>java testBM
Kromoson : ABCDEFG
Cek Kromoson : ABCDEFG
Tidak Terjadi Mutasi Kromosom
Jumlah Langkah : 6
  
```

Table 5 Tidak Mutasi - Boyer Moore

- Kategori Mutasi Inversi – KMP

```

Command Prompt
F:\Kuliah\StiMa\Tugas Besar III\Setor>java test
Kromoson : ABCDEFG
Cek Kromoson : AECDBFG
Terjadi Mutasi Kromosom Inversi
Jumlah Langkah : 10
  
```

Table 6 Mutasi Inversi-KMP

- Kategori Mutasi Inversi – Boyer Moore

```

Command Prompt
F:\Kuliah\StiMa\Tugas Besar III\Setor>java testBM
Kromoson : ABCDEFG
Cek Kromoson : AECDBFG
Terjadi Mutasi Kromosom Inversi
Jumlah Langkah : 2
  
```

Table 7 Mutasi Inversi - Boyer-Moore

- Kategori Duplikasi dengan Boyer-Moore

```

Command Prompt
F:\Kuliah\StiMa\Tugas Besar III\Setor>java testBM
Kromoson : ABCDEFG
Cek Kromoson : ABBCEDEFG
Mutasi Kromosom Duplikasi
  
```

Table 8 Kategori Duplikasi

- Kategori Delesi dengan KMP

```

Command Prompt
F:\Kuliah\StiMa\Tugas Besar III\Setor>java test
Kromoson : ABCDEFG
Cek Kromoson : ACDEFG
Mutasi Kromosom Delesi
  
```

Table 9 Katgori Mutasi Delesi

IV. ANALISIS

Untuk kategori tidak terjadi mutasi dan Inversi digunakan pencocokan string seperti biasa dengan KMP maupun boyer-moore. Untuk kategori Delesi akan dicek jika panjang string kromosom mutan lebih pendek dari kromosom induk maka dinyatakan bahwa kromosom tersebut delesi sebelum memasuki keadaan untuk mencocokkan *string*. Begitu pula dengan kategori Duplikasi yang akan dicek jika panjang *string* kromosom mutan lebih panjang dari kromosom tidak termutasi, maka dapat disimpulkan bahwa mutasi kromosom tersebut dalam kategori duplikasi.

Dari hasil uji dapat dilihat perbandingan dari KMP dan Boyer-Moore dimana Boyer-Moore menghasilkan langkah yang lebih minimum tinimbang KMP.

V KESIMPULAN

Mutasi kromosom adalah perubahan materi genetic yang disebabkan oleh perubahan pada kromosom. Mutasi kromosom dapat disebabkan oleh gangguan fisik dan kimia yang menyebabkan kesalahan di dalam pembelahan sel sehingga merusak susunan kromosom atau mengubah jumlah kromosom. Perubahan susunan kromosom tersebut dapat disebabkan oleh **delesi, duplikasi, inversi, dan translokasi**.

Dari algoritma *pattern matching* yang dicoba algoritma Boyer-Moore lebih mangkus tinimbang algoritma KMP hal ini terlihat dari jumlah langkah yang dihasilkan.

VI. UCAPAN TERIMA KASIH

Penulis ingin mengucapkan terima kasih kepada Bapak Rinaldi dan Ibu Ulfah selaku dosen Strategi Algoritma (IF2211) yang telah mengajarkan dasar-dasar teori yang penulis butuhkan untuk menyelesaikan makalah ini dan telah membimbing penulis selama satu semester ini. Penulis juga ingin mengucapkan terima kasih kepada teman-teman seperjuangan dan keluarga atas dukungannya selama ini.

REFERENSI

- [1] Munir, Rinaldi. Strategi Algoritms. Informatika Bandung, 2009
- [2] Aryulina, Diah dkk. BIOLOGI SMA kelas XII , Esis Bandung 2007
- [3] <http://softilmu.blogspot.com/2014/09/pengertian-dan-macam-macam-mutasi.html>

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 4 Mei 2015



Mochamad Try Yulyanto (13513067)