

Penggunaan Algoritma *String Matching* untuk Mendeteksi Kanker

Ivana Clairine Irsan - 13512041
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganessa 10 Bandung 40132, Indonesia
ivanaclairine@itb.ac.id

Tujuan algoritma *string matching* adalah untuk mendapatkan kemunculan suatu pattern pada sebuah text. Teknik ini juga telah digunakan pada bidang analisis DNA, salah satunya untuk menentukan apakah DNA seseorang telah bermutasi dan beresiko mengidap penyakit kanker atau tidak. Karena memegang peranan yang penting, algoritma yang digunakan terus berkembang, dan pada makalah ini akan dibahas peranan algoritma *string matching* dalam pencocokan DNA untuk mendeteksi kanker dan beberapa contohnya.

Index Terms—kanker, DNA, *string matching*.

I. Pendahuluan

Kanker adalah penyakit mematikan yang hingga saat ini banyak menyebabkan kematian di negara berkembang, tepatnya kanker menyebabkan 25% kematian di seluruh dunia dan dalam setahun, sekitar 0.5% populasi terkena kanker. Kebanyakan kanker dapat dirawat dan disembuhkan, terutama bila perawatan dimulai sejak awal. Kesulitan untuk menyembuhkannya semakin sulit jika saat ditemukan kanker sudah berkembang hingga tahap akhir. Karena keganasannya inilah penyakit kanker banyak menyita banyak perhatian para peneliti karena penyakit ini sering tidak terdeteksi hingga mencapai stadium akhir. Namun, bukanlah hal yang mudah untuk mendeteksi penyakit ini karena kanker terdiri dari banyak jenis seperti kanker prostat, kanker paru-paru, kanker kolon dan rektum, kanker darah, kanker kulit, dll., dan banyak orang tidak menyadari dirinya menderita kanker hingga suatu saat kanker yang dideritanya sudah mencapai stadium akhir. Saat kanker sudah mencapai stadium akhir adalah saat di mana biasanya penderita mulai menunjukkan gejala penyakit secara fisik. Hal inilah yang harus dicegah karena saat kanker sudah memasuki tahap akhir, sangat besar kemungkinan kanker sudah menyebar ke organ tubuh yang lain dan proses pengobatan menjadi semakin sulit dilakukan.

Banyak usaha telah dilakukan oleh dokter dan peneliti untuk menemukan metode yang paling tepat untuk mendeteksi bibit kanker sejak berada pada tahap awal, bahkan saat kanker masih menjadi sebuah kelainan DNA dan belum berkembang menjadi penyakit. Salah satu cara yang dianggap efektif adalah dengan melakukan tes darah

karena darah mengandung DNA yang menjadi ‘identitas’ dan ‘pengenal’ seseorang. Teknik yang paling sering digunakan untuk pencocokan DNA adalah *string matching*.

String matching adalah teknik pencocokan string yang dapat digunakan pada berbagai bidang. Pada kehidupan sehari-hari pun, sebagian besar orang sudah sering menggunakan implementasi dari algoritma string matching, misalnya seperti fungsi “*find*” pada *Microsoft Word* dan pencarian *file* dalam komputer. Walaupun algoritma *string matching* banyak digunakan pada kehidupan sehari-hari untuk melakukan banyak hal kecil, ternyata algoritma ini juga mempunyai peranan yang sangat besar dalam kesejahteraan hidup manusia karena ia memegang peranan penting dalam dunia medis.

Pada makalah ini akan ditunjukkan bagaimana peranan string matching pada pencarian bibit kanker pada rantai DNA dengan menggunakan algoritma Boyer-Moore dan KMP.

II. Dasar Teori

2.1 DNA

Di dalam tubuh manusia, tepatnya di dalam inti sel, terdapat sebuah rantai asam deoksiribonukelat, atau yang biasa disebut DNA. Setiap aktivitas sel akan tersimpan dan dapat dideteksi melalui DNA karena ia berperan sebagai cetak biru aktivitas sel dalam tubuh makhluk hidup.

DNA pertama kali berhasil diambil dari dalam nukleus (inti sel) pada tahun 1868 oleh ilmuwan berkebangsaan Swiss Friedrich Miescher di Jerman. Walaupun DNA telah didapatkan sejak tahun 1868, penelitian terhadap peranan DNA di dalam sel baru dimulai pada awal abad 20.

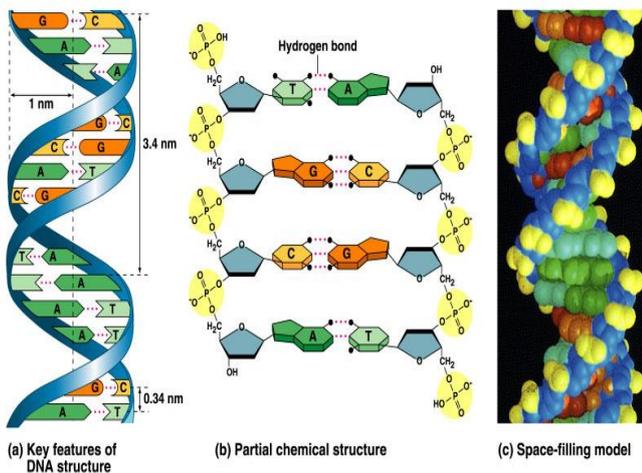
DNA merupakan polimer yang terdiri dari 3 komponen utama, yaitu :

- gugus fosfat
- gula deoksiribosa
- basa nitrogen, yang terdiri dari 4 jenis :
 - o Adenin (A)
 - o Guanin (G)
 - o Sitosin (C)
 - o Timin (T)

Sebuah unit monomer DNA yang terdiri dari ketiga komponen di atas dinamakan nukleotida, sehingga DNA disebut sebagai polinukleotida.

Rantai DNA memiliki lebar 22-24 Å, sementara panjang satu unit nukleotida 3,3 Å. Walaupun unit monomer ini sangatlah kecil, DNA dapat memiliki jutaan nukleotida yang terangkai seperti rantai. Misalnya, kromosom terbesar pada manusia terdiri atas 220 juta nukleotida. Rangka utama untai DNA dibentuk oleh gugus fosfat dan gula yang berselang seling. Gula pada DNA disebut gula pentosa.

DNA terdiri atas dua untai yang berpilin membentuk struktur heliks ganda. Masing-masing untai terdiri atas rangka utama dan basa nitrogen. Kedua untai pada heliks ganda DNA disatukan oleh ikatan hydrogen Antara basa-basa yang terdapat pada kedua untai tersebut. Empat basa yang ditemukan pada DNA (Adenin(A), Guanin(G), Sitosin(C), dan Timin(T)) masing-masing akan berikatan dengan pasangannya.



Gambar 1 : Struktur DNA

Hubungan antar basa nitrogen, seperti yang bisa dilihat pada Gambar 1 adalah sebagai berikut :

- Adenin(A) >> Timin(T)
- Guanin(G) >> Sitosin (C)

Segmen polipeptida dari DNA disebut gen, yang biasanya merupakan sebuah bentuk dari molekul RNA. Berbeda dengan DNA, struktur gula yang membentuk RNA adalah ribosa.

2.1.1 Sekuens DNA

Informasi biologis seseorang didapatkan dari menganalisis bagian dari makhluk hidup yang membawa informasi genetik, dan informasi ini berada di dalam DNA yang membawa informasi genetik tiap makhluk hidup.

Karena DNA merupakan ciri kunci makhluk hidup, pengetahuan mengenai sekuens DNA dapat berguna dalam setiap penelitian

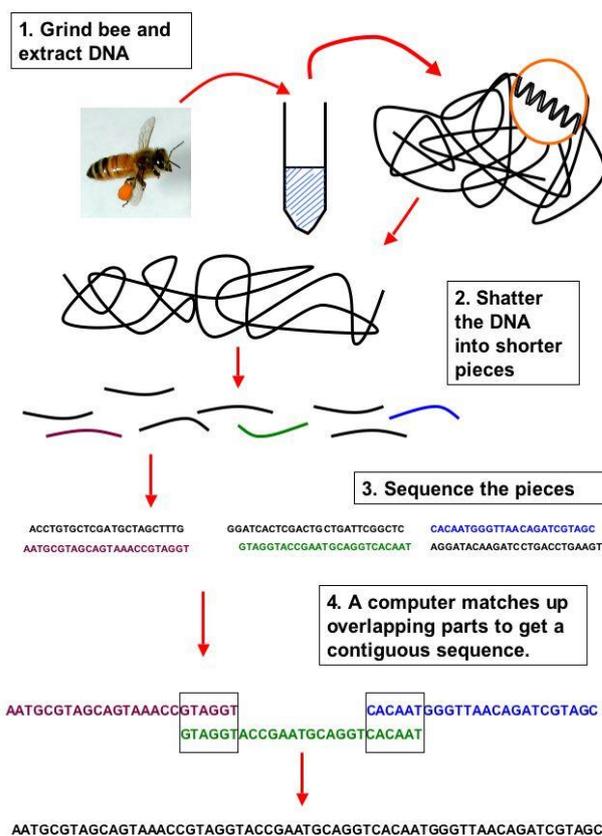
Sekuens DNA adalah sebuah proses untuk mendapatkan susunan lengkap nukleotida rantai DNA dari suatu makhluk hidup.

Sekuens DNA menyandikan informasi yang diperlukan bagi makhluk hidup untuk melangsungkan hidup dan berkembang biak. Pengetahuan akan sekuens DNA berguna untuk mengetahui sekuens asam amino yang disandikan oleh gen.

Pada mulanya, sekuensing DNA dilakukan dengan mentranskripsikannya ke dalam bentuk RNA terlebih dahulu karena metode sekuensing RNA telah ditemukan sebelumnya.

Saat ini sekuens DNA sangat berguna dalam bidang kedokteran karena dapat digunakan untuk mengidentifikasi, mendiagnosis, dan mengembangkan pengobatan terhadap penyakit genetik.

Hampir semua sampel biologis dapat memberikan sebuah salinan DNA yang dapat digunakan dalam proses sekuens DNA. Dengan sejumlah kecil air liur, rambut, sel epitel, darah, dan bagian tubuh manapun yang mengandung DNA, bisa dilakukan sekuens DNA.



Gambar 2 : Proses sekuens DNA untuk mendapatkan susunan nukleotida lebah

Ada beberapa metode yang digunakan dalam proses sekuens DNA, di antaranya adalah :

- Metode Maxam-Gilbert

Metode ini awalnya cukup populer karena dapat langsung menggunakan DNA hasil pemurnian, sedangkan metode Sanger pada waktu itu memerlukan kloning untuk membentuk DNA untai tunggal. Sayangnya, kerumintan pengerjaan teknis yang harus dilakukan saat menggunakan metode Maxam-Gilbert yang ini sangat

rumit dan kesulitannya untuk dibuat dalam skala besar membuat metode ini mulai ditinggalkan.

- *Metode Sanger (Terminasi Rantai)*

Saat ini, hampir semua pengerjaan sekuensing DNA dilakukan dengan menggunakan metode terminasi rantai yang dikembangkan oleh Frederick Sanger dan rekan-rekannya. Teknik ini menggunakan terminasi atau penghentian reaksi sintesis DNA menggunakan substrat nukleotida yang telah dimodifikasi.

Seiring dengan perkembangannya, kini terdapat beberapa macam metode sekuensing terminasi rantai yang berbeda satu sama lain terutama dalam hal pendeteksian fragmen DNA hasil reaksi sekuensing.

2.2 Kanker

Kanker adalah penyakit yang ditandai dengan kelainan siklus sel yang menyebabkan sel mampu tumbuh tidak terkendali, menyerang jaringan biologis di dekatnya, dan bermigrasi ke jaringan tubuh yang lain melalui sirkulasi darah atau sistem limfatik.

Cabang ilmu kedokteran yang berhubungan dengan studi, diagnosis, perawatan, dan pencegahan kanker disebut onkologi.

Pada umumnya, kanker dibedakan berdasarkan jenis organ atau sel tempat ditemukannya sel yang sudah terserang kanker. Kanker yang bermula di liver akan disebut sebagai kanker hati, dan kanker yang bermula di usus akan disebut kanker usus. Klasifikasi kanker kemudian dilakukan pada kategori yang lebih umum, yaitu :

- **Karsinoma**

Karsinoma merupakan kanker yang terjadi pada jaringan epitel seperti kulit atau jaringan yang menyelubungi organ tubuh, misalnya organ pada sistem pencernaan atau kelenjar.

Contoh : kanker kulit, karsinoma serviks, karsinoma anal, karsinoma esofageal, karsinoma hepatoselular, kanker laryngeal, hipernefroma, kanker lambung, kanker testikular, dan kanker tiroid.

- **Sarkoma**

Sarkoma merupakan kanker yang terjadi pada tulang

Contoh : osteosarkoma, kondrosarkoma (tulang rawan), rabdomiosarkoma (jaringan otot)

- **Leukemia**

Leukemia merupakan kanker yang terjadi akibat tidak matangnya sel darah yang berkembang di dalam sumsum tulang dan memiliki kecenderungan untuk berakumulasi di dalam sirkulasi darah.

- **Limfoma**

Limfoma merupakan kanker yang timbul dari nodus limfa dan jaringan dalam sistem kekebalan tubuh

Ada 3 kondisi yang dapat menyebabkan perubahan sel normal menjadi sel kanker, yaitu :

- **Hiperplasia**

Hiperplasia adalah keadaan saat sel normal dalam jaringan bertumbuh dalam jumlah yang berlebihan.

- **Displasia**

Displasia merupakan kondisi ketika sel berkembang tidak normal dan pada umumnya terlihat adanya

perubahan pada nukleusnya. Pada tahapan ini ukuran nukleus bervariasi, aktivitas mitosis meningkat, dan tidak ada ciri khas sitoplasma yang berhubungan dengan diferensiasi sel pada jaringan

- **Neoplasia**

Neoplasia merupakan kondisi sel pada jaringan yang sudah berproliferasi secara tidak normal dan memiliki sifat invasive

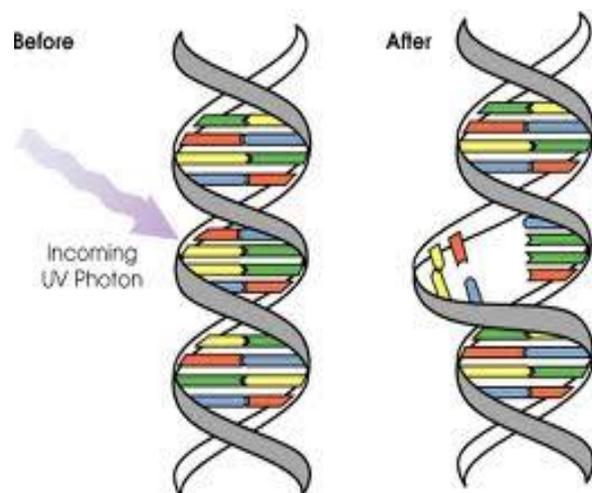
Pertumbuhan yang tidak terkenali dari sel tersebut disebabkan karena adanya kerusakan DNA yang menyebabkan mutasi di gen vital yang mengontrol pembelahan sel. Beberapa mutasi dapat mengubah sel normal menjadi sel kanker. Mutasi-mutasi ini sering diakibatkan agen kimia maupun fisik yang disebut karsinogen.

Kanker merupakan penyakit mematikan yang bertanggung jawab atas 25% kematian di dunia. Pada tahun 2012, tercatat sebesar 8.2 juta kematian disebabkan oleh penyakit ini. Di Indonesia sendiri, saat ini diperkirakan pertumbuhan kanker adalah sebesar 170-190 kasus setiap 100000 jiwa, dan saat ini kanker berada di peringkat 3 penyebab kematian terbesar di Indonesia setelah penyakit jantung.

Sayangnya, sebagian besar penderita kanker (60-70%) baru mendapatkan perawatan di saat penyakit mereka sudah mencapai stadium akhir sehingga sudah sangat terlambat untuk dilakukan perawatan.

Kanker yang paling banyak menyebabkan kematian di seluruh dunia per tahunnya adalah kanker paru-paru, kanker hati, kanker usus, dan kanker payudara. Lebih dari 60% dari total penderita kanker berada di Afrika, Asia, Amerika Tengah, dan Amerika Selatan. Daerah-daerah tersebut berkontribusi atas 70% dari seluruh jumlah kematian yang disebabkan oleh kanker di dunia.

WHO memperkirakan, jumlah kasus kematian yang disebabkan oleh kanker akan meningkat dari 14 juta jiwa menjadi 22 juta jiwa dalam 20 tahun mendatang.



Gambar 3 : Ilustrasi mutasi gen yang disebabkan oleh radiasi sinar UV dan jika tidak ditangani sesegera mungkin dapat menyebabkan kanker

Sumber : <http://kirandhillon.com.s94760.gridserver.com/>

2.2.1 Teknik Diagnosis Kanker

Saat ini sudah banyak praktek dan penelitian mengenai cara mendeteksi kanker yang dilakukan oleh para pelaku dunia medis dan peneliti. Meskipun banyak teknik yang telah ditemukan dan banyak yang sedang dikembangkan, masih belum ditemukan teknik yang dapat memberikan hasil yang 100% akurat, yang disebabkan oleh terbatasnya teknologi yang digunakan, keterbatasan mesin, hingga kesalahan manusia (*human error*).

Teknik yang saat ini sering digunakan untuk mendiagnosa kanker adalah biopsi, endoskopi, tes darah, tes Pap, dan tes genetic.

Biopsi adalah sebuah prosedur untuk mengambil sampel sel dari organ tubuh untuk kemudian dianalisis di dalam laboratorium. Biasanya, sebelum biopsi akan dilakukan *imaging test* (X-Ray, rontgen) yang akan memberi gambaran organ tubuh mana yang abnormal (tidak berfungsi dengan baik). Karena imaging test hanya memberikan hasil berupa gambar, sulit untuk menentukan apakah sel yang terdapat di organ tersebut terjerangkit kanker atau tidak sehingga perlu dilakukan biopsi.

Cara mendeteksi kanker selain biopsi yang telah diteliti selama beberapa tahun terakhir adalah metode tes genetik (*Genetic Testing*).

Tes genetik adalah proses membaca kode DNA untuk mengidentifikasi adanya bagian yang tidak normal. Hal ini didukung oleh perkembangan teknologi sekuens DNA yang membawa perkembangan dalam ilmu komputasi DNA.

Salah satu teknik tes genetic yang paling terkenal adalah FISH (*Fluorescence in Situ Hybridization*), yang tekniknya adalah meletakkan specimen di bawah sinar fluorescent dan dibandingkan dengan bagian dari gen yang disebut probe. Warna yang dihasilkan oleh sinar fluorescent akan mendeteksi bagian mana dari DNA yang mengalami mutasi (penggalan DNA merupakan rantai tunggal yang didapatkan sebagai hasil dari sekuens DNA).

2.2 Pencocokan String (*String Matching*)

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 adalah :

- Algoritma *Brute Force* (*Naïve Algorithm*)

Brute Force akan mencocokkan string di setiap karakter untuk menentukan apakah pattern yang dimaksud terdapat di posisi tersebut

Dengan algoritma Brute Force, pattern akan maju 1 langkah ke kanan dan mulai mencocokkan lagi sampai bertemu dengan karakter yang tidak cocok, pattern yang dimaksud sudah ditemukan, atau pencarian sudah mencapai ujung teks.

Kompleksitas waktu terburuk: $O(MN)$

Kompleksitas waktu terbaik: $O(N)$

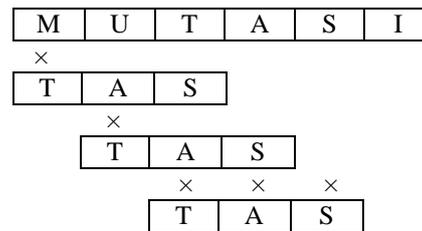
Pseudocode :

```
do
if (text letter == pattern letter)
compare next letter of pattern to next
letter of text
else
move pattern down text by one letter
while (entire pattern found or end of
text)
```

Contoh :

Teks = "MUTASI"

Pattern = "TAS"



Jumlah perbandingan : 5

- Algoritma Knuth-Morris-Pratt (KMP)
Algoritma Knuth-Morris-Pratt (KMP) berbeda dengan algoritma Brute Force karena algoritma ini menyimpan informasi dari hasil perbandingan yang dilakukan sebelumnya untuk menghindari perbandingan yang sia-sia. KMP menggunakan menggunakan prefix dan suffix dari pattern untuk mengoptimasi pergeseran pattern dalam pencarian.
Kompleksitas waktu total : $O(n + m)$

Contoh :

Teks : AGAAATTGC

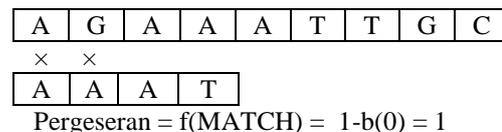
Pattern : AAAT

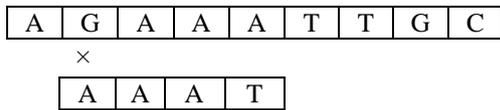
Tabel Fungsi Pembatas:

p	0	1	2
b(p)	0	1	2

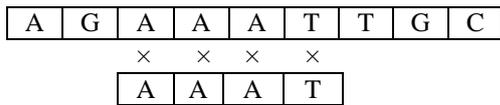
$$f(\text{MATCH}) = \text{MATCH} - b(\text{MATCH}-1) \text{ untuk } \text{MATCH} \geq 1$$

$$f(\text{MATCH}) = 1 \text{ untuk } \text{MATCH} = 0$$





Pergeseran = $f(\text{MATCH}) = 1$



Jumlah perbandingan : 7

Pseudocode :

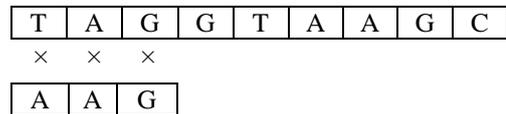
```

KMPMatch(T,P)
Input: Strings T (text) with n
characters and P
(pattern) with m characters.
Output: Starting index of the first
substring of T
matching P, or an indication that P is
not a
substring of T.
f ← KMPFailureFunction(P) {build
failure function}
i ← 0
j ← 0
while i < n do
if P[j] = T[i] then
if j = m - 1 then
return i - m - 1 {a match}
i ← i + 1
j ← j + 1
else if j > 0 then {no match, but we
have advanced}
j ← f(j-1) {j indexes just after
matching prefix in P}
else
i ← i + 1
return "There is no substring of T
matching P"

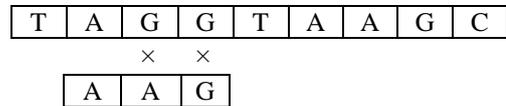
```

- *Algoritma Boyer-Moore*
Algoritma pencocokan string Boyer Moore didasarkan pada 2 teknik, yaitu :
 1. *Looking-glass technique*
Teknik ini merupakan cara untuk menemukan sebuah pattern dalam teks dengan memulai pencocokan dari akhir string pattern.
 2. *Character-jump technique*
Saat terjadi ketidakcocokan, pencarian akan dilanjutkan setelah menggeser pattern sebesar nilai tertentu untuk menghindari pencocokan yang sia-sia

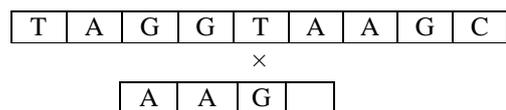
Contoh :
Teks : TAAGTAAGC
Pattern : AAG



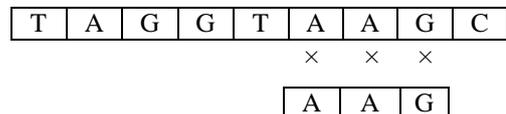
Pergeseran = 1, karena pattern "AAG" tidak mengandung huruf "T" sehingga pattern dimajukan sebanyak 1 indeks dengan letak indeks pertama pattern terletak di indeks teks+1



Pergeseran = 1, karena pattern mengandung huruf "G" tapi sudah terlewatkan



Pergeseran = 3, karena pattern "AAG" tidak mengandung huruf "T" sehingga pattern dimajukan sebanyak 1 indeks dengan letak indeks pertama pattern terletak di indeks teks+1



Jumlah perbandingan : 9

Pseudocode :

```

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 - 1
Return "no match"

```

III. Hasil Pendeteksian Rantai DNA yang Bermutasi

Hasil pendeteksian adanya DNA yang bermutasi (dengan algoritma Boyer-Moore):

```

C:\Windows\system32\cmd.exe
D:\ITB\Semester4\STIMA\MAKALAH>java DNABM
String DNA (sampel yang akan dites): ggtagtgaatgattaaacttgcctc
ctagtggtgcccagtaggaacgacaacagcagctgagggcttaaacattaatgagttctcat
gcggaatagaactgtgagggcgtggaacaaactctttgaccgcggctctttactaccac
gtctgagggagtagagccgttggagtcgatttggatggtctgacgacagccatctcgc
ccatcttgagatttggctcaatttggatgacctaaactcgttgaatgagtcagttccagag
aatctccaattatgagttcttatcatctgacaggaacatgaagtcagcagcggcaatccc
tggaaactcttaaacacagagaacccagccagcgaacatgggtgccatgctgcagcctagacc
aggacaccgctatgcaaatcttacccagcaatccgaagccggtgtaaatgggtgctctag
cgttttctcaaggagctcgtcttaactgactctccgataaccaatgtgcataggctgcggtggg
totgtggggaagaaccggcgaccgctgactgattaaactcagcatcgatggcactgg
accaatccaagatttttctgctaggtgagtcactcgtggccggatgaggaagcgggtgta
cacatactggctaaagtcctcaggtcgtgtaaccaagtcacattctctgcaaatgactagg
atctggctgaaatccgagctgtctgtaaggaaggaaggagatattcactgaaagccctt
gcgtaccattcttagatattaggaagctgactaggggataattgctcaaatctaacggg
tatatgttagggctcactaacctgttagcctgactatgtagatccagaccctcgggaaacc
tggagcctctatcatcaatacactacagaatctgctcgggtgttctaaatttgcga
String DNA normal :ggtagtgaatgattaaacttgcctcctagtgggtgcccaggtgaa
ccgacaaccagctgagggctcaaacatattgtagtctctatcggaatagaactgtgagcc
ggtgaaaccatactctttgcaacgggtcttttactaccagctcaggggagtaacagcc
ggttggagtcgatttggatggtctgacgagaccgcatctcgcactctgagatttggccca
tttctgatacacttaattcgttgaatgagtgagttcagagaatcctccaattatgctct
ttatcatctgacaggaacatgaagtcagcagcggcaatcccggaaactcttaagcatag
agaaccagccagcgaacattgctgctcactgctgcaactagaccaggaaccgctattgcaaa
tttgcaccagtaatccgaagccggtgtaatagggtagcgtctagcgtttcccgaagagctc
tcttaactgactctccgataccaatggtgcataggctcgggtgggtcttgggggaaagaaccg
cgaccgtacagctgattatgcatccagcatcgatcgatactgaaccaatccaagatttatct
tcgctaggtgagtaactcgtggtccgagtaggaagcgggtttacacatactggctaaagctc
catgctcgggttaaccaagctcactctctcgaactgactaggatctggctgcaatccgagc
tgtctgtaaggaagagaagagattatcactgaagccctcgtgactaccattcttagatc
caggaagctgactagggatattatgctcaaatctaacgcctacagataggggtactata
actgttagcctgactctgtagatccagaccgctcgggaaagctggagcctctatcatcata
atacactacagaatcgcgtcgggtgttctaaatttgcga
Jumlah perbandingan : 65
Terdeteksi adanya mutasi pada DNA
    
```

Hasil pendeteksian adanya DNA yang bermutasi (dengan algoritma Knuth-Morris-Pratt):

```

C:\Windows\system32\cmd.exe
D:\ITB\Semester4\STIMA\MAKALAH>javac DNAKMP.java
D:\ITB\Semester4\STIMA\MAKALAH>java DNAKMP
String DNA (sampel yang akan dites): ggtagtgaatgattaaacttgcctc
ctagtggtgcccagtaggaacgacaacagcagctgagggcttaaacattaatgagttctcat
gcggaatagaactgtgagggcgtggaacaaactctttgaccgcggctctttactaccac
gtctgagggagtagagccgttggagtcgatttggatggtctgacgacagccatctcgc
ccatcttgagatttggctcaatttggatgacctaaactcgttgaatgagtcagttccagag
aatctccaattatgagttcttatcatctgacaggaacatgaagtcagcagcggcaatccc
tggaaactcttaaacacagagaacccagccagcgaacatgggtgccatgctgcagcctagacc
aggacaccgctatgcaaatcttacccagcaatccgaagccggtgtaaatgggtgctctag
cgttttctcaaggagctcgtcttaactgactctccgataccaatgtgcataggctgcggtggg
totgtggggaagaaccggcgaccgctgactgattaaactcagcatcgatggcactgg
accaatccaagatttttctgctaggtgagtcactcgtggccggatgaggaagcgggtgta
cacatactggctaaagtcctcaggtcgtgtaaccaagtcacattctctgcaaatgactagg
atctggctgaaatccgagctgtctgtaaggaaggaaggagattatcactgaaagccctt
gcgtaccattcttagatattaggaagctgactaggggataattgctcaaatctaacggg
tatatgttagggctcactaacctgttagcctgactatgtagatccagaccctcgggaaacc
tggagcctctatcatcaatacactacagaatctgctcgggtgttctaaatttgcga
String DNA normal :ggtagtgaatgattaaacttgcctcctagtgggtgcccaggtgaa
ccgacaaccagctgagggctcaaacatattgtagtctctatcggaatagaactgtgagcc
ggtgaaaccatactctttgcaacgggtcttttactaccagctcaggggagtaacagcc
ggttggagtcgatttggatggtctgacgagaccgcatctcgcactctgagatttggccca
tttctgatacacttaattcgttgaatgagtgagttcagagaatcctccaattatgctct
ttatcatctgacaggaacatgaagtcagcagcggcaatcccggaaactcttaagcatag
agaaccagccagcgaacattgctgctcactgctgcaactagaccaggaaccgctattgcaaa
tttgcaccagtaatccgaagccggtgtaatagggtagcgtctagcgtttcccgaagagctc
tcttaactgactctccgataccaatggtgcataggctcgggtgggtcttgggggaaagaaccg
cgaccgtacagctgattatgcatccagcatcgatcgatactgaaccaatccaagatttatct
tcgctaggtgagtaactcgtggtccgagtaggaagcgggtttacacatactggctaaagctc
catgctcgggttaaccaagctcactctctcgaactgactaggatctggctgcaatccgagc
tgtctgtaaggaagagaagagattatcactgaagccctcgtgactaccattcttagatc
caggaagctgactagggatattatgctcaaatctaacgcctacagataggggtactata
actgttagcctgactctgtagatccagaccgctcgggaaagctggagcctctatcatcata
atacactacagaatcgcgtcgggtgttctaaatttgcga
Jumlah perbandingan : 1219
Terdeteksi adanya mutasi pada DNA
    
```

IV. Analisis Hasil Pencocokan DNA

Lakukan sekuens DNA untuk mendapatkan rantai tunggal yang akan diperiksa
 Contoh (penggalan dari string DNA yang digunakan pada bab III):
 ggta gtga atga ttaa aact tgcc tcct
 agtg ggtg ccag gtag gaac

Pada string yang normal, string pasangannya (sesuai dengan hubungan antar basa) **seharusnya** adalah sebagai berikut:
 ccat cact tact aatt ttga acgg agga
 tcac ccac ggtc catc **C**ttg

Namun, pada DNA yang telah bermutasi (tidak normal), akan terbentuk rantai yang tidak konsisten dengan rantai pasangannya. Pada contoh di atas, string pasangan yang terlihat adalah:
 ccat cact tact aatt ttga acgg agga
 tcac ccac ggtc catc **t**gg

Saat akan dicocokkan dengan menggunakan string matching, yang dicocokkan adalah converse dari pasangan yang seharusnya, pada contoh di atas, String normal sebagai teks.

Teks: "ggta gtga atga ttaa aact tgcc tcct agtg ggtg ccag gtag gaac"

Hasil converse dari DNA yang didapat dari pasien (mungkin tidak normal) sebagai pattern.
 String DNA pasien:

"ggta gtga atga ttaa aact tgcc tcct agtg ggtg ccag gtag caac"

Hasil converse (digunakan sebagai Pattern):
 "ggta gtga atga ttaa aact tgcc tcct agtg ggtg ccag gtag aaac"

Dengan algoritma Boyer Moore dan KMP, akan ditunjukkan bahwa makhluk hidup pemilik DNA yang sedang diuji memiliki kemungkinan terjangkit kanker karena adanya rangkaian DNA yang rusak pada bagian berikut:

NORMAL:
 "ggta gtga atga ttaa aact tgcc tcct agtg ggtg ccag gtag **aaac**"

PASIEEN:
 "ggta gtga atga ttaa aact tgcc tcct agtg ggtg ccag gtag **caac**"

Dari hasil di atas, dapat dilihat bahwa ada keabnormalitasan DNA pasien. Jika tidak segera diteliti dan ditangani, mutasi DNA yang terjadi ini dapat berkembang menjadi kanker yang berujung pada kematian.

Dari segi algoritma string matching, karena adanya perbedaan teknik perbandingan dari algoritma KMP dan Boyer-Moore, jumlah perbandingan untuk teks dan pattern yang sama juga berbeda.

Dapat dilihat dalam bab III, bahwa jumlah perbandingan untuk pencocokan DNA dengan panjang string 1000 membutuhkan 65 perbandingan, sedangkan jumlah perbandingan saat menggunakan algoritma KMP dengan string yang sama membutuhkan 1219 kali perbandingan.

V. Kesimpulan

DNA yang membawa informasi mengenai kondisi tubuh makhluk hidup dapat digunakan untuk mendeteksi apakah seseorang sedang terkena penyakit. Hal ini dapat dilakukan dengan menggunakan peralatan yang mengimplementasikan algoritma *string matching*. Algoritma string matching akan mengidentifikasi apakah DNA sampel yang diteliti mengalami mutasi atau tidak sehingga dapat dilakukan penanganan bagi penderita jika memang diketahui DNA penderita telah mengalami mutasi dan perlu diperbaiki.

Dari dua algoritma yang telah dicoba, dapat dilihat bahwa algoritma boyer-moore lebih efektif dalam melakukan pencocokan DNA karena jumlah perbandingan yang dilakukan lebih sedikit dibandingkan dengan jumlah perbandingan yang dilakukan oleh algoritma Knuth-Morris-Pratt.

VI. Ucapan Terima Kasih

Penulis ingin mengucapkan kepada Tuhan karena telah memampukan penulis untuk menyelesaikan makalah ini dengan baik tepat waktu, juga kepada Bapak Rinaldi dan Ibu Masayu yang telah mengajarkan dasar-dasar teori yang penulis butuhkan untuk menyelesaikan makalah ini. Tidak hanya bahan untuk makalah ini, penulis juga ingin mengucapkan terima kasih kepada kedua dosen yang telah membimbing dan menemani kami selama satu semester ini mempelajari semua pelajaran Strategi Algoritma yang menarik dan telah memberikan pengetahuan baru yang berharga bagi penulis.

References

- [1] "The DNA sequence and biological annotation of human chromosome 1". *Nature* **441** (7091): 315–21
- [2] Munir, Rinaldi. *Strategi Algoritma*. Informatika Bandung, 2009
- [3] <http://www.cancerresearchuk.org/cancer-info/cancerandresearch/all-about-cancer/what-is-cancer/making-new-cells/dna-damage-and-repair/DNA-damage-repair> -- diakses pada tanggal 13 Mei 2014
- [4] <http://www.banggangan.com/wallpaper/36073/dna-double-helix.html> -- diakses pada tanggal 14 Mei 2014
- [5] <http://www.wisegeek.org/> -- diakses pada tanggal 15 Mei 2014
- [6] http://jjco.oxfordjournals.org/content/32/suppl_1/S17.full -- diakses pada tanggal 16 Mei 2014
- [7] <http://www.who.int/mediacentre/factsheets/fs297/en/> -- diakses pada tanggal 16 Mei 2014

- [8] http://www.bioinformatics.org/sms2/sample_dna.html -- diakses pada tanggal 16 Mei 2014
- [9] http://www.bioinformatics.org/sms2/mutate_dna.html -- diakses pada tanggal 16 Mei 2014
- [10] http://www.bioinformatics.org/sms2/rev_comp.html -- diakses pada tanggal 17 Mei 2014

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, 17 Mei 2014



Ivana Clairine Irsan
13512041