

Algoritma *String Matching* yang Efektif untuk Mendeteksi Penuaan

Freddy Isman - 13513007
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13513007@std.stei.itb.ac.id

Algoritma *string matching* bertujuan untuk mencari kemunculan *pattern* pada sebuah teks. Teknik ini juga digunakan pada analisis DNA, yaitu untuk mendeteksi apakah DNA seseorang mengalami mutasi atau kerusakan sehingga mengakibatkan penuaan yang dialaminya. Pengembangan algoritma ini cukup luas karena peranannya yang cukup penting dan pada makalah ini akan dibahas mengenai peranan algoritma *string matching* dalam pencocokan DNA untuk mengidentifikasi lokasi mutasi atau kerusakan gen yang menyebabkan penuaan.

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

I. PENDAHULUAN

Penuaan merupakan kondisi kesehatan yang tidak dapat dihindari oleh setiap makhluk hidup, termasuk manusia. Penuaan akan bertambah tinggi seiring dengan bertambahnya usia. Meskipun begitu, kecepatan penuaan di berbagai bagian tubuh tidaklah sama. Antara tahun 2000 dan tahun 2050, proporsi penduduk di atas 60 tahun akan bertambah dua kali lipat dari 11% menjadi 22%. Jumlah penduduk di atas 60 tahun diperkirakan akan meningkat dari 605 juta menjadi 2 miliar di periode yang sama. Data di atas menandakan bahwa penduduk yang mengalami penuaan akan bertambah seiring dengan waktu. Salah satu dampak yang paling dikenal dari penuaan adalah penurunan fungsi organ tubuh, yang mana akan menurunkan kinerja dan mengganggu aktivitas sehari-hari. Penuaan bukanlah suatu penyakit, namun keberadaannya akan menurunkan kondisi kesehatan makhluk hidup yang kemudian akan membuatnya rentan untuk terkena penyakit yang lain.

Usaha telah banyak dilakukan oleh dokter dan para ahli terkait serta peneliti untuk menemukan metode yang paling tepat dalam mendeteksi kerusakan atau mutasi pada gen DNA. Salah satu cara yang dianggap efektif yaitu dengan

melakukan tes sampel darah karena di dalam sampel darah seseorang terdapat kandungan DNA yang menjadi identitas atau ciri khas seseorang. Kemudian DNA yang diperoleh akan dicocokkan dengan menggunakan teknik *string matching*.

String matching adalah teknik pencocokan string yang dapat digunakan pada berbagai jenis bidang. Penggunaan terbesarnya yaitu dalam bidang komputer, seperti fungsi “*find*” pada *Microsoft Word* dan pencarian *file* pada komputer. Walaupun begitu, tidak menutup kemungkinan bagi algoritma *string matching* ini untuk dipakai dalam bidang medis, salah satunya yang paling penting yaitu dalam kaitannya dengan pencocokan DNA.

Pada makalah ini penulis akan membahas bagaimana peranan *string matching* pada pendeteksian mutasi atau kerusakan gen pada DNA yang menjadi penyebab penuaan dengan menggunakan algoritma Boyer-Moore dan Knuth-Morris-Pratt.

II. DASAR TEORI

I. 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.

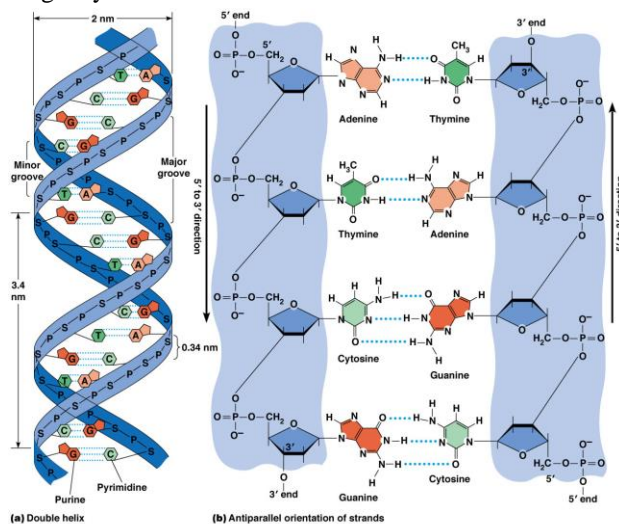
DNA pertama kali berhasil diambil dari dalam nukleus (inti sel) pada tahun 1868 oleh ilmuwan berkebangsaan Swiss Friedrich Miescher di Jerman. Walaupun DNA telah didapatkan sejak tahun 1868, penelitian terhadap peranan DNA di dalam sel baru dimulai pada awal abad 20.

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

- gugus fosfat

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

Sebuah unit monomer DNA yang terdiri dari ketiga komponendi atas dinamakan nukleotida, sehingga DNA disebut sebagai polinukleotida. Rantai DNA memiliki lebar 22-24Å, sementara panjang satu unit nukleotida 3,3 Å. Walaupun unit monomerini sangatlah kecil, DNA dapat memiliki jutaan nukleotida yang terangkai seperti rantai. Misalnya, kromosom terbesar pada manusia terdiri atas 220 juta nukleotida. Rangka utama untaian DNA dibentuk oleh gugus fosfat dan gula yang berselang seling. Gula pada DNA disebut gula pentosa. DNA terdiri atas dua untaian yang berpilin membentuk struktur heliks ganda. Masing-masing untaian terdiri atas rangka utama dan basa nitrogen. Kedua untaian pada heliks ganda DNA disatukan oleh ikatan hydrogen Antara basa-basa yang terdapat pada kedua untaian 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.

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 kuncimakhluk 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 aminoyang 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.

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, kerumitan 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 Penuaan

Penuaan (aging) adalah suatu proses menghilangnya secara perlahan-lahan kemampuan jaringan untuk

memperbaiki diri dan mempertahankan struktur dan fungsi normalnya sehingga tidak dapat memperbaiki kekurangan yang diderita.

Tidak seorangpun yang dapat menghentikan proses penuaan. Siklus ini ditandai dengan tahap-tahap mulai menurunnya berbagai fungsi organ tubuh karena setelah mencapai dewasa, secara alamiah seluruh komponen tubuh tidak dapat berkembang lagi. Sebaliknya justru terjadi penurunan karena proses penuaan. Penuaan merupakan suatu proses multidimensional, yang tidak hanya terkait dengan faktor jasmani, tapi juga psikologis dan sosial. Penuaan itu sendiri adalah suatu proses alamiah kompleks yang melibatkan setiap molekul, sel dan organ dalam tubuh.

Menua (menjadi tua) adalah suatu proses menghilangnya secara perlahan-lahan kemampuan jaringan untuk memperbaiki diri/mengganti dan mempertahankan fungsi normalnya sehingga tidak dapat bertahan terhadap infeksi dan memperbaiki kerusakan yang diderita. (Constantinescu, 1994)

Proses menua bukan merupakan suatu penyakit, melainkan suatu masa atau tahap hidup manusia yaitu: bayi, kanak-kanak, dewasa, tua, dan lanjut usia. Orang mati bukan karena lanjut usia tetapi karena suatu penyakit, atau juga suatu kecacatan. Akan tetapi proses menua dapat menyebabkan berkurangnya daya tahan tubuh dalam menghadapi rangsangan dari dalam maupun luar tubuh. Walaupun demikian, memang harus diakui bahwa ada berbagai penyakit yang sering menghinggapi kaum lanjut usia.

Proses menua sudah mulai berlangsung sejak seseorang mencapai usia dewasa. Misalnya dengan terjadinya kehilangan jaringan pada otot, susunan saraf, dan jaringan lain sehingga tubuh mati sedikit demi sedikit. Sebenarnya tidak ada batas yang tegas, pada usia berapa penampilan seseorang mulai menurun. Pada setiap orang, fungsi fisiologis alat tubuhnya sangat berbeda, baik dalam hal pencapaian puncak maupun menurunnya.

Proses Tahapan Penuaan

Penuaan tidak terjadi secara tiba-tiba, melainkan melalui beberapa tahapan atau fase, sehingga kita memiliki kesempatan untuk menghambatnya, salah satunya dengan menjaga pola makan dan pemakaian krim atau pelembab untuk melindungi kulit dari sengatan matahari agar kulit tidak cepat kering atau keriput. Proses penuaan terjadi

secara bertahap dan secara garis besar dapat dibagi menjadi 3 fase:

Fase 1 Subklinik

Pada saat mencapai usia 25-35 tahun. Pada masa ini produksi hormon mulai berkurang (mulai mengalami penurunan produksi). Pada tahap ini, sebagian besar hormon di dalam tubuh mulai menurun, yaitu hormon testosteron, growth hormon, dan hormon estrogen. Pembentukan radikal bebas, yang dapat merusak sel dan DNA, mulai memengaruhi tubuh. Polusi udara, diet yang tak sehat dan stres merupakan serangan radikal bebas yang dapat merusak sel-sel tubuh. Kerusakan ini biasanya tak tampak dari luar. Karena itu, pada tahap ini orang merasa dan tampak normal, tidak mengalami gejala dan tanda penuaan. Di fase ini mulai terjadi kerusakan sel tapi tidak memberi pengaruh pada kesehatan. Tubuh pun masih bugar terus. Penurunan ini mencapai 14 % ketika seseorang berusia 35 tahun.

Fase 2 Transisi

Kedua transisi, yakni pada usia 35-45 tahun. Produksi hormon sudah menurun sebanyak 25%, sehingga tubuh pun mulai mengalami penuaan. Biasanya pada masa ini, ditandai dengan lemahnya penglihatan (mata mulai mengalami rabun dekat) sehingga perlu menggunakan kacamata berlensa plus, rambut mulai beruban, stamina dan energi tubuh pun berkurang. Bila pada masa ini dan sebelumnya atau bila pada usia muda, kita melakukan gaya hidup yang tidak sehat bisa berisiko terkena kanker.

Fase 3 Klinik

Puncaknya pada tahap fase klinikal, yakni pada usia 45 tahun ke atas. Pada masa ini produksi hormon sudah berkurang hingga akhirnya berhenti sama sekali. Kaum perempuan mengalami masa yang disebut menopause sedangkan kaum pria mengalami masa andropause. Pada masa ini kulit pun menjadi kering karena mengalami dehidrasi/kulit menjadi keriput, terutama di bagian samping dan di bawah mata kita, juga kulit tangan kita yang tidak sekenyal dulu, tubuh juga menjadi cepat lelah. Berbagai penyakit degeneratif seperti diabetes, osteoporosis, hipertensi dan penyakit jantung koroner mulai menyerang dan menjadi sesuatu yang sangat mengerikan.

Karena proses penuaan ini terjadi melalui beberapa

tahapan, sebenarnya ada banyak waktu untuk menghambatnya. Cepat lambatnya proses penuaan, 30% dipengaruhi oleh faktor genetika/keturunan dan 70 % lebih dipengaruhi oleh gaya hidup. Kalau anggota keluarga cenderung awet muda. Kita pun besar kemungkinan akan berpenampilan awet muda. Gaya hidup yang penuh stres, kurang istirahat, banyak makan makanan berlemak dan berkalori tinggi, kurang gerak serta hidup di lingkungan yang penuh polusi akan merusak sel sehingga menjadi lebih tua. Akibatnya, kita pun mengalami penuaan usia biologik. Namun, kondisi ini dapat dihindari dengan program anti aging baik yang dilakukan sendiri maupun dengan bantuan medis. Misalnya: Seseorang yang rajin berolahraga, terbukti bisa menangkai sejumlah penyakit kardiovaskuler. Olah raga ringan di sela aktivitas seperti senam, lari atau jalan cepat sebaiknya sering dilakukan.

emakin jauh seseorang dari derita penyakit jantung, stroke dan sejenisnya, Semakin berbahagia hidupnya. Dan kebahagiaan itu merupakan salah satu peran terbesar penunda penuaan. Tidak mungkin rasanya orang bisa terlihat sehat dan awet muda kalau tubuhnya dihindangi berbagai jenis penyakit berbahaya. Penunda penuaan lainnya adalah faktor diet dan nutrisi. Apa yang kita makan menentukan tubuh kita. Diet dan nutrisi sangat berperan dalam menentukan proses penuaan dan kesehatan seseorang.

2.3 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"

M U T A S I

×

T A S

×

T A S

× × ×

T A S

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

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

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

A G A A A T T G C

× ×

A A A T

Pergeseran = $f(\text{MATCH}) = 1 - b(0) = 1$

A G A A A T T G C

× ×

A A A T

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

A G A A A T T G C

× ×

A A A T

Jumlah perbandingan : 7

-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

T A G G T A A G C

× × ×

A A G

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

T A G G T A A G C

× ×

A A G

Pergeseran = 1, karena pattern mengandung huruf “G” tapi sudah melewati

T A G G T A A G C

×

A A G

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

T A G G T A A G C

× × ×

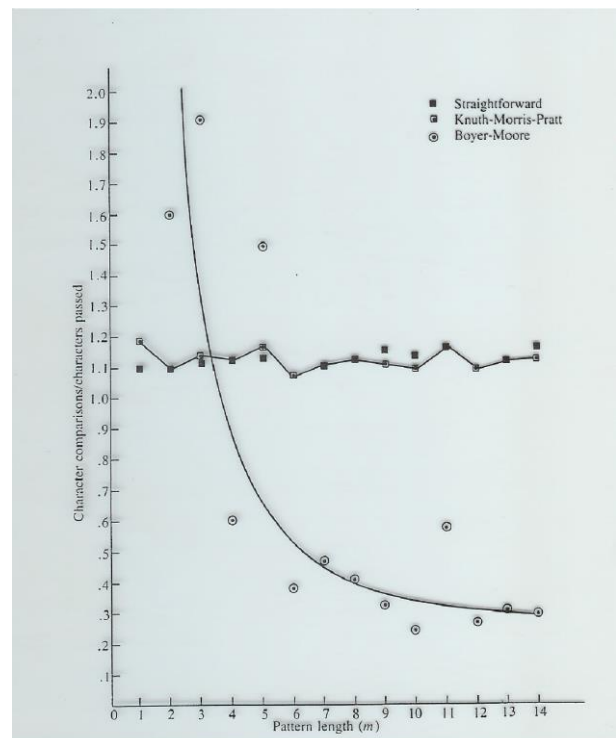
A A G

Jumlah perbandingan : 9

III. ANALISIS EFEKTIFITAS ALGORITMA STRING MATCHING PADA PENCOCOKAN DNA

Dari dasar teori, diketahui bahwa jumlah macam kodon yang ada di rantai DNA cukup bervariasi dan karena adanya hubungan basa kodon yang telah dijelaskan sebelumnya, maka pada rantai DNA tidak dimungkinkan adanya *long uniform suffiks* (suffiks panjang yang sama/seragam, misalnya “baaaaaaaaaaaaaaa” (suffiks adalah bagian yang digaris bawah)), sehingga algoritma Boyer-Moore tidak akan mencapai kasus terburuknya (*worst-case*).

Grafik hasil analisis terhadap masing-masing algoritma string matching pada pencocokan DNA diberikan sebagai berikut:



Dari grafik di atas, dapat kita analisis bahwa efektifitas performansi algoritma Boyer-Moore bertambah seiring dengan bertambahnya panjang pattern atau rantai gen yang dicocokkan. Hal ini ditunjukkan dengan mengecilnya nilai (jumlah karakter yang dibandingkan/jumlah karakter yang dilewati) selama proses pencocokan berlangsung. Namun, untuk algoritma lainnya, seperti algoritma Straight-Forward/Brute-Force dan algoritma Knuth-Morris-Pratt, kedua algoritma ini tidak menunjukkan peningkatan efektifitas performansi seiring dengan bertambahnya panjang pattern atau rantai gen yang dicocokkan. Dan untuk panjang pattern ≥ 3 , kedua algoritma ini memiliki performansi yang lebih buruk jika dibandingkan dengan algoritma Boyer-Moore. Karena panjang pattern atau rantai gen yang dibandingkan atau

dicocokkan jauh lebih besar, maka dapat dilihat bahwa algoritma Boyer-Moore jauh lebih efektif dalam melakukan *string matching* pada rantai gen DNA.

IV. KESIMPULAN DAN SARAN

DNA yang membawa informasi mengenai karakteristik dari makhluk hidup dapat dideteksi mutasi atau kerusakan pada rantai gennya yang menyebabkan terjadinya penuaan pada makhluk hidup sehingga untuk kedepannya dapat dilakukan penanganan untuk mengatasi penuaan dengan memperbaiki gen yang termutasi atau yang mengalami kerusakan. Hal ini dapat dilakukan dengan mengimplementasikan algoritma *string matching*.

Algoritma *string matching* akan mengidentifikasi gen pada sampel DNA yang diteliti dengan menggunakan dua macam algoritma, yaitu algoritma Boyer-Moore dan algoritma Knuth-Morris-Pratt. Dari dua algoritma tersebut, penulis menyarankan penggunaan algoritma Boyer-Moore yang lebih efektif dalam melakukan pencocokan DNA karena jumlah perbandingan yang dilakukan lebih sedikit dibandingkan dengan jumlah perbandingan yang dilakukan oleh algoritma Knuth-Morris-Pratt.

V. UCAPAN TERIMA KASIH

Penulis ingin mengucapkan kepada Tuhan karena telah memberikan kemampuan bagi penulis untuk menyelesaikan makalah ini dengan baik tepat waktu, juga kepada Bapak Rinaldi Munir dan Ibu Nur Ulfa Maulidevi 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 penulis dan yang lain selama satu semester ini mempelajari semua pelajaran Strategi Algoritma yang menarik dan telah memberikan pengetahuan baru yang berharga bagi penulis.

REFERENSI

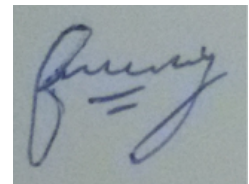
- [1] Munir,Rinaldi."Pencocokan string (*string matching/pattern matching*)", diakses pada tanggal 4 Mei 2015.
- [2] <http://www.who.int/ageing/about/facts/en/> -- diakses pada tanggal 4 Mei 2015.
- [3] <http://people.cs.pitt.edu/~kirk/cs1501/Pruhs/Spring2006/notes/> -- diakses pada tanggal 4 Mei 2015.

- [4] <http://www.bangngangan.com/wallpaper/36073/dna-double-helix.html> -- diakses pada tanggal 4 Mei 2015.
- [5] <http://www.wisegEEK.org/> -- diakses pada tanggal 4 Mei 2015.
- [6] "The DNA sequence and biological annotation of human chromosome 1".Nature 441(7091): 315–21
- [7] http://www.bioinformatics.org/sms2/sample_dna.html -- diakses pada tanggal 4 Mei 2015.
- [8] http://www.bioinformatics.org/sms2/mutate_dna.html -- diakses pada tanggal 4 Mei 2015.
- [9] http://www.bioinformatics.org/sms2/rev_comp.html -- diakses pada tanggal 4 Mei 2015.
- [10] <http://www.cse.iitd.ac.in/~mohanty/col106/Resources/Rabin-Karp.ppt> -- diakses pada tanggal 4 Mei 2015.

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



Freddy Isman - 13513007