

# Pendeteksian Penyakit Malaria melalui Uji DNA dengan Algoritma Pencocokan String KMP

Hafis Alrafi Irsal, 13516034

Program Studi Teknik Informatika

Sekolah Teknik Elektro dan Informatika

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

hafisalrafi@gmail.com

**Abstract**—Malaria merupakan salah satu penyakit menular yang disebabkan oleh parasit tipe *Plasmodium* sp. Parasit ini akan masuk ke dalam darah seseorang melalui perantara gigitan nyamuk *Anopheles* betina yang terinfeksi. Parasit penyebab penyakit malaria masuk ke dalam genus *Plasmodium* dan salah satu spesies yang menginfeksi manusia adalah *Plasmodium falciparum*. Parasit jenis termasuk yang paling berbahaya karena menyerang semua bentuk eritrosit (sel darah merah). Saat ini telah banyak diteliti metode untuk mendeteksi penyakit malaria, salah satunya adalah dengan uji keberadaan DNA dari *Plasmodium* pada DNA sampel (orang yang kemungkinan mengidap malaria). Deteksi ini dilakukan dengan penggunaan algoritma pencocokan string. Pencocokan DNA dari *Plasmodium* dan sampel ini mirip dengan konsep pencocokan string yang biasa dipakai pada bidang informatika. Pada pencocokan DNA ini kita mencari kecocokan antara DNA pattern (dalam hal ini DNA *Plasmodium*) dengan DNA sampel (orang yang ingin diidentifikasi apakah mengidap penyakit malaria). DNA terdiri dari satuan-satuan molekul yang disebut nukleotida. Tiap nukleotida terdiri atas gula monosakarida (deoksiribosa), gugus fosfat, dan salah satu jenis basa nitrogen Guanin (G), Adenin (A), Timin (T), atau Sitosin (C). Karena DNA mengandung basa nitrogen tersebut maka DNA bisa digambarkan sebagai kumpulan gugus karbon yang dapat dianalogikan sebagai karakter. Sehingga DNA bisa diibaratkan sebagai kumpulan karakter atau string. Maka dari itu, algoritma yang digunakan pada pencocokan string dapat diterapkan untuk pencocokan DNA. Untuk melakukan pengujian DNA *Plasmodium* dan sampel, pada makalah ini akan memakai salah satu algoritma yang mangkus dalam pencocokan string yaitu algoritma Knuth-Morris-Pratt (KMP).

**Kata kunci**—Malaria; DNA; *Plasmodium*; Algoritma KMP

## I. PENDAHULUAN

Malaria adalah penyakit yang ditularkan melalui perantara nyamuk *Anopheles* yang terinfeksi oleh parasit tipe *Plasmodium* sp. Malaria menyebabkan gejala yang biasanya termasuk demam, kelelahan, muntah, dan sakit kepala. Dalam kasus yang parah dapat menyebabkan kulit kuning, kejang, koma, atau kematian. Sebagian besar kematian pada pengidap

penyakit malaria ini disebabkan oleh *P. falciparum* karena menyerang semua bentuk eritrosit (sel darah merah). Sementara *P. vivax*, *P. ovale*, and *P. malariae* umumnya menyebabkan bentuk yang lebih ringan.

Menurut data WHO, sekitar 4,2 miliar orang (hampir separuh populasi dunia) rentan terserang malaria. Pada 2015, ditemukan 214 juta orang terinfeksi malaria dan 438.000 di antaranya meninggal. Angka kematian akibat malaria jauh lebih tinggi dibandingkan dengan ebola yang memicu kematian 4000 orang di Afrika tahun 2015 lalu.

Banyak metode yang dilakukan untuk mendiagnosis penyakit ini, diantaranya dengan pemeriksaan mikroskopis darah menggunakan film darah, atau dengan uji diagnostik cepat berdasarkan-antigen. Di samping itu, ada cara pendeteksian penyakit malaria dengan pengujian DNA *Plasmodium* dan sampel dengan menggunakan instrumen reaksi rantai berpolimerase. Tetapi metode ini belum banyak digunakan karena biaya dan kompleksitasnya. Pada makalah ini, akan dibahas bagaimana cara kerja pencocokan DNA *Plasmodium* dan sampel (penderita penyakit malaria) dengan menggunakan algoritma pencocokan string KMP.

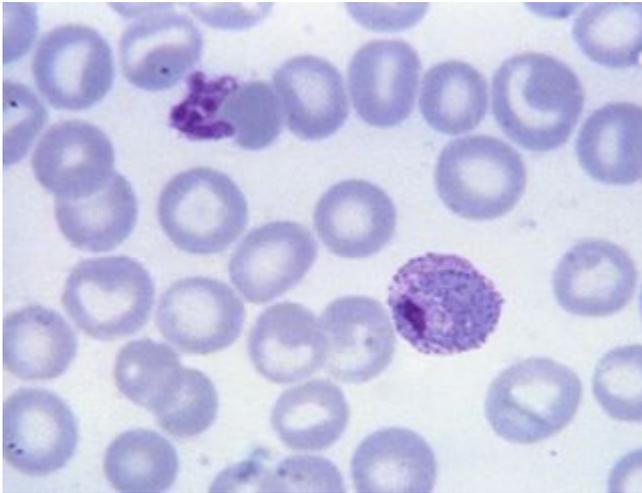
## II. KAJIAN TEORI

### 2.1. Malaria

Malaria merupakan penyakit menular yang disebabkan oleh parasit *Plasmodium* sp. Malaria menyebar melalui gigitan nyamuk *Anopheles* yang sudah terinfeksi oleh parasit. Malaria akan mengakibatkan kematian jika tidak ditangani dengan benar. Infeksi malaria bisa terjadi hanya dengan satu gigitan nyamuk. Malaria jarang sekali menular secara langsung dari satu orang ke orang lainnya. Contoh kondisi penularan penyakit ini adalah jika terjadi kontak dengan darah penderita atau janin bisa terinfeksi karena tertular dari darah sang ibu.

Parasit selanjutnya menyerang eritrosit (sel darah merah) dan mulai memakan hemoglobin. Pecahnya sel darah merah

yang terinfeksi plasmodium ini dapat menyebabkan timbu disertai menggigil. Karena banyak sel darah merah yang pecah, maka menyebabkan anemia.

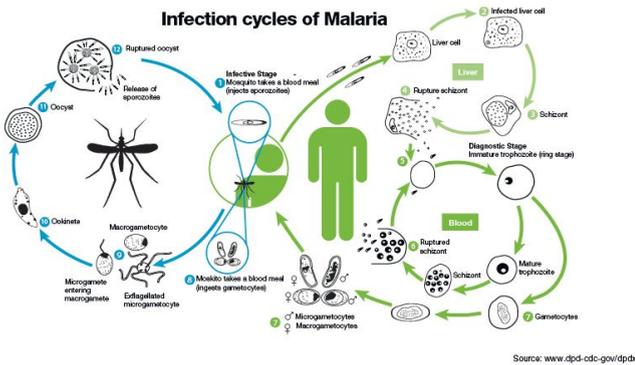


Gambar 1. Sel darah merah yang terserah parasit *Plasmodium sp.*

(Sumber: <http://malariana.blogspot.co.id>)

Penyebab Penyakit Malaria adalah Plasmodium, terdapat beberapa jenis plasmodium yaitu plasmodium vivax, plasmodium ovale, plasmodium falciparum dan malaria plasmodium. Khusus untuk plasmodium falciparum sering menjurus ke malaria berat yang sering menyebabkan kemarian. Sedangkan jenis plasmodium lainnya menyebabkan malaria skala ringan dan jarang menjurus ke malaria akut.

2.2. Plasmodium



Gambar 2. Siklus Malaria yang disebabkan parasit *Plasmodium sp.*

(Sumber: <http://pharma.bayer.com>)

Plasmodium merupakan genus protozoa parasit. Penyakit yang disebabkan oleh genus ini dikenal sebagai malaria. Parasit ini sentiasa mempunyai dua inang dalam siklus

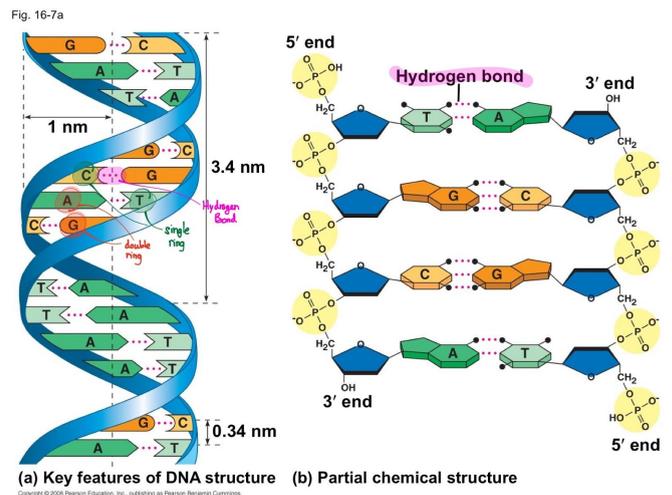
hidupnya: vektor nyamuk dan inang vertebra. Sekurang-kurangnya sepuluh spesies menjangkiti manusia. Spesies lain menjangkiti hewan lain, termasuk burung, reptilia dan hewan pengerat. Beberapa jenis plasmodium yang menyerang manusia yaitu plasmodium vivax, plasmodium ovale, plasmodium falciparum dan malaria plasmodium. Plasmodium falciparum merupakan parasit yang paling berbahaya karena menyerang semua bentuk eritrosit (sel darah merah).

2.3. DNA

DNA merupakan suatu asam nukleat yang menyimpan segala informasi biologis dan instruksi-instruksi genetika yang unik dari setiap makhluk hidup. Instruksi-instruksi genetika ini berperan penting dalam pertumbuhan, perkembangan, dan fungsi organisme. Molekul DNA terdiri dari dua unting biopolimer yang berpilin satu sama lainnya membentuk heliks ganda (double helix). Dua unting DNA ini dikenal sebagai polinukleotida karena keduanya terdiri dari satuan-satuan molekul yang disebut nukleotida. Tiap nukleotida terdiri atas gula monosakarida (deoksiribosa), gugus fosfat, dan salah satu jenis basa nitrogen. Basa nitrogen yang ada pada DNA berupa

- Guanin (G),
- Adenin (A),
- Timin (T), atau
- Sitosin (C).

Nukleotida-nukelotida ini kemudian tersambung dalam satu rantai ikatan kovalen antara gula satu nukleotida dengan fosfat nukelotida lainnya. Hasilnya adalah rantai punggung gula-fosfat yang berselang-seling. Menurut kaidah pasangan basa (A dengan T dan C dengan G), ikatan hidrogen mengikat basa-basa dari kedua unting polinukleotida membentuk DNA unting ganda.



Gambar 3. Struktur DNA

(Sumber: <http://logyofbio.blogspot.co.id>)

### III. ALGORITMA KNUTH-MORRIS-PRATT (KMP)

Algoritma KMP dikembangkan oleh D. E. Knuth, bersama-sama dengan J. H. Morris, dan V. R. Pratt. Jika pada algoritma pencocokan string dengan brute force, setiap kali ada ketidakcocokan pattern dengan teks, maka pattern digeser satu karakter ke kanan, maka 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.

#### 3.1. Fungsi Pinggiran (Border Function)

Pada algoritma Knuth-Morris-Pratt (KMP) akan melakukan proses awal (preprocessing) terhadap pattern P dengan menghitung fungsi pinggiran (beberapa literatur menyebut fungsi overlap, fungsi failure, fungsi awalan) yang mengindikasikan pergeseran s terbesar yang mungkin dengan menggunakan perbandingan yang dibentuk sebelum pencarian string. Dengan adanya fungsi pinggiran ini, dapat dicegah pergeseran yang sia-sia seperti halnya pada algoritma brute force.

Fungsi pinggiran hanya bergantung pada karakter-karakter di dalam pattern, dan bukan pada karakter-karakter di dalam teks yang dicari. Oleh karena itu, kita dapat melakukan perhitungan fungsi awalan sebelum pencarian string dilakukan.

Algoritma untuk menghitung fungsi pinggiran.

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] }
```

Deklarasi

k,q : integer

Algoritma:

```
b[1] <- 0
q <- 2
k <- 0
for q <- 2 to m do
```

```
while ((k > 0) and (P[q] != P[k+1])) do
    kb <- [k]
endwhile
if P[q]=P[k+1] then
    k <- k+1
endif
b[q]=k
endfor
```

Fungsi pinggiran b(j) didefinisikan sebagai ukuran awalan terpanjang dari P yang merupakan akhiran dari P[1..j].

Sebagai contoh, tinjau pattern P = abacab. Nilai F untuk setiap karakter di dalam P adalah sebagai berikut.

j	0	1	2	3	4	5
P[j]	a	b	a	c	a	b
k	-	0	1	2	3	4
b(k)	-	0	0	1	0	1

Tabel 1. Fungsi pinggiran dari pattern "abacab"

Pada tabel fungsi pinggiran di atas dapat dilihat nilai dari b(4) = 1. Nilai b(4) didapati dengan cara berikut. Cari prefix yang paling panjang dari P[0..4] yang juga suffix dari P[1..4]. Berarti kita mencari prefix paling panjang dari "abaca" yang juga suffix dari "baca".Maka ditemukan "a" yang mempunyai panjang 1. Sehingga nilai b(4) adalah 1.

#### 3.2. Perbandingan (Pencocokan) String

Langkah selanjutnya adalah melakukan pencocokan string dengan menggunakan algoritma Knuth-Morris-Pratt sebagai berikut.

Langkah-langkah yang dilakukan algoritma Knuth-Morris-Pratt pada saat mencocokkan string.

1. Algoritma Knuth-Morris-Pratt mulai mencocokkan pattern pada awal teks.
2. Dari kiri ke kanan, algoritma ini akan mencocokkan karakter per karakter pattern dengan karakter di teks yang bersesuaian, sampai salah satu kondisi berikut dipenuhi:
  - a. Karakter di pattern dan di teks yang dibandingkan tidak cocok (mismatch).

- b. Semua karakter di pattern cocok. Kemudian algoritma akan memberitahukan penemuan di posisi ini.
3. Algoritma kemudian menggeser pattern berdasarkan tabel fungsi pinggiran, lalu mengulangi langkah 2 sampai pattern berada di ujung teks.

**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)
```

```
{ Mencari kecocokan pattern P di dalam
teks T dengan algoritma
Knuth-Morris-Pratt. Jika ditemukan P di
dalam T, lokasi awal kecocokan disimpan
di dalam peubah idx.
```

```
Masukan: pattern P yang panjangnya m
dan teks T yang panjangnya n.
Teks T direpresentasika
sebagai string (array of character)
```

```
Keluaran: posisi awal kecocokan
(idx). Jika P tidak ditemukan, idx = -
1.
}
```

**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
    jb[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
```

**3.3. Kompleksitas**

Pada algoritma KMP mempunyai fungsi pinggiran. Untuk menghitung fungsi pinggiran dibutuhkan waktu  $O(m)$ , sedangkan pencarian string membutuhkan waktu  $O(n)$ , sehingga kompleksitas waktu algoritma KMP adalah  $O(m+n)$ .

**IV. PENERAPAN ALGORITMA KMP DALAM PENCOCOKAN DNA**

DNA dapat dianalogikan sebagai rangkaian karakter yang terdiri atas karakter A, C, T, G. Oleh sebab itu, pencocokan DNA dapat dianalogikan sebagai pencocokan string. Algoritma KMP yang akan dipakai dalam pencocokan DNA sangat mangkus jika karakter alphabet kecil dan seringkali terjadi pengulangan karakter seperti pada untaian DNA.

Berikut ini contoh cara kerja pencocokan DNA pattern (P) dengan sampel (S).

DNA Sampel : AGTGAGTAGTACAGTAGCGTATTT

DNA Pattern : AGTAGC

Langkah pertama yang dilakukan adalah mencari fungsi pinggiran.

j	0	1	2	3	4	5
P[j]	A	G	T	A	G	C
k	-	0	1	2	3	4
b(k)	-	0	0	0	1	2

Tabel 2. Fungsi Pinggiran dari DNA pattern "AGTAGC"

Pertama kita melakukan pengecekan dari awal.

S : AGTGAGTAGTACAGTAGCGTATTT

P : AGTAGC

Karakter 1-3 sama, tetapi pada posisi ke-4 karakter S dan P berbeda. Maka, kita lakukan pergeseran sesuai dengan fungsi pinggiran. Karakter yang berbeda adalah A indeks ke-3 pada pattern, maka indeks dari pattern dikurangi dengan nilai b(k) nya. Maka kita dapatkan  $3-0=3$ . Jadi, pattern akan digeser sejauh 3 karakter.

S : AGTGAGTAGTACAGTAGCGTATTT

P : AGTAGC

Kemudian kita bandingkan lagi mulai dari karakter ke-4. Lalu langsung ditemukan ketidakcocokan maka dilakukan pergeseran sebesar 1 langkah sesuai dengan fungsi pinggiran.

S : AGTGAGTAGTACAGTAGCGTATTT

P : AGTAGC

Lanjut kita membandingkan mulai dari karakter ke-5. Maka ditemukan ketidakcocokan pada S dan P di karakter ke-10.

S : AGTGAGTAGTACAGTAGCGTATTT

P : AGTAGC

Kita lanjutkan pencocokan mulai dari karakter ke-8 dan ditemukan ketidakcocokan pada karakter ke-12. Maka lakukan pergeseran sejauh 3 karakter sesuai dengan fungsi pinggiran.

S : AGTGAGTAGTACAGTAGCGTATTT

P : AGTAGC

Mulai lagi pengecekan dari karakter ke-11 dan ditemukan ketidakcocokan pada karakter ke-12. Maka dilakukan pergeseran sejauh 1 karakter.

S : AGTGAGTAGTACAGTAGCGTATTT

P : AGTAGC

Kembali dilakukan pengecekan dimulai dari karakter ke-12 dan langsung ditemukan ketidakcocokan pada karakter ke-12 tersebut. Lalu lakukan pergeseran 1 karakter.

S : AGTGAGTAGTACAGTAGCGTATTT

P : AGTAGC

Kita lanjutkan pencocokan DNA mulai dari karakter ke-13 dan hingga ujung pattern tidak ditemukan lagi ketidakcocokan antara pattern dan sampel, maka pencocokan berhasil dan DNA pattern ditemukan dalam DNA sampel.

## V. ANALISIS PENDETEKSIAN PENYAKIT MALARIA DENGAN PENCOCOKAN DNA

Dari pengaplikasian cara kerja pencocokan DNA di atas, dapat kita lihat bahwa untuk melakukan pendeteksian apakah seorang menderita penyakit malaria adalah melalui kecocokan DNA pattern yaitu DNA dari parasit *Plasmodium sp.* dan DNA sampel yaitu seorang yang akan dideteksi menderita penyakit malaria.

Ketika pada pencocokan DNA, ditemukannya DNA pattern (parasit malaria) dalam DNA sampel, maka dapat ditentukan bahwa orang tersebut menderita penyakit malaria.

Namun, untuk melakukan pencocokan DNA sampel dan pattern dalam kasus pendeteksian malaria dibutuhkan instrumen lain dari bidang medis, misal berupa polimerase chain reaction (PCR) untuk mengecek DNA Plasmodium pada sampel. Penggunaan instrumen ini masih belum banyak digunakan karena biaya dan tingkat kompleksitasnya.

Meskipun keberadaan DNA Plasmodium mengindikasikan adanya infeksi parasit malaria, dibutuhkan studi lebih lanjut untuk membandingkan hasil yang didapatkan dengan menggunakan pemeriksaan lainnya.

## VI. KESIMPULAN

Algoritma pencocokan String Knuth-Morris-Pratt (KMP) sangat cocok dan mangkus untuk digunakan dalam algoritma untuk pencocokan DNA. DNA yang hanya terdiri dari 4 jenis gugus karbon dianalogikan sebagai sebuah string, maka DNA tersebut merupakan kombinasi dari 4 jenis karakter. Dengan demikian sangat dimungkinkan bahwa akan terjadi perulangan karakter dalam DNA. Karena akan seringnya terjadi perulangan karakter itu, maka Algoritma pencocokan String Knuth-Morris-Pratt (KMP) sangat cocok untuk digunakan sebagai Algoritma pencocokan DNA.

Memfaatkan metode pencocokan DNA tersebut dapat dideteksi penyakit malaria yang mengidap seseorang dengan mencocokkan DNA sampel dengan DNA parasit malaria. Jika pada DNA sampel ditemukan DNA pattern (parasit), maka dapat dikatakan seseorang tersebut positif mengidap penyakit malaria. Namun, studi lebih lanjut terhadap metode ini sangat diharapkan agar memberikan hasil dengan tingkat akurasi tinggi dan biaya yang lebih murah. Di zaman serba teknologi ini, penggunaan Algoritma Knuth-Morris-Pratt (KMP) akan bisa menjadi alternatif dalam pendeteksian penyakit malaria melalui metode pencocokkan DNA.

## UCAPAN TERIMA KASIH

Penulis ingin mengucapkan Alhamdulillah, segala puji bagi Allah yang telah memberikan kemampuan untuk menyelesaikan makalah ini. Penulis juga ingin menyampaikan ucapan terima kasih kepada Ibu Masayu selaku dosen mata studi Strategi Algoritma yang telah mengajarkan dasar-dasar teori yang diperlukan penulis untuk menyelesaikan makalah ini. Terakhir, penulis juga mengucapkan terima kasih kepada setiap orang yang terlibat dalam penulisan makalah ini.

## REFERENSI

- [1] Rinaldi Munir, Diktat kuliah IF2211 Strategi Algoritma, Teknik Informatika ITB, 2018.
- [2] Anany Levitin, Introduction to the Design & Analysis of Algorithms, Addison-Wesley, 2003.
- [3] <http://dinkes.inhukab.go.id/?p=4028>, diakses tanggal 10 Mei 2018.

## 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, 12 Mei 2018



Hafis Alrafi Irsal  
13516034