

**IMPLEMENTASI ALGORITMA
NEEDLEMAN-WUNSCH DAN SMITH-WATERMAN
PADA PENJAJARAN SEKUEN VIRUS ZIKA**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer (S.Kom.)**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Jason Anggada

12110110003

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2016

LEMBAR PENGESAHAN SKRIPSI

**IMPLEMENTASI ALGORITMA
NEEDLEMAN-WUNSCH DAN SMITH-WATERMAN
PADA PENJAJARAN SEKUEN VIRUS ZIKA**

Oleh

Nama : Jason Anggada

NIM : 12110110003

Program Studi : Teknik Informatika

Fakultas : Teknik dan Informatika

Tangerang, 12 Agustus 2016

Ketua Sidang

Dosen Penguji

Maria Irmina Prasetyowati, S.Kom., M.T. Yustinus Widya Wiratama, S.Kom., M.Sc.

Dosen Pembimbing I

Dosen Pembimbing II

Ranny, S.Kom., M.Kom.

Eka Antonius Kurniawan, S.Kom., M.Sc.

Mengetahui,
Ketua Program Studi
Teknik Informatika

Maria Irmina Prasetyowati, S.Kom., M.T.

PERNYATAAN TIDAK MELAKUKAN PLAGIAT

Dengan ini saya,

Nama : Jason Anggada
NIM : 121110110003
Program Studi : Teknik Informatika
Fakultas : Teknik dan Informatika

menyatakan bahwa skripsi yang berjudul **“Implementasi Algoritma Needleman-Wunsch dan Smith-Waterman pada Penjajaran Sekuen Virus Zika”** ini adalah karya ilmiah saya sendiri, bukan plagiat dari karya ilmiah orang lain atau lembaga lain, dan semua karya ilmiah orang lain atau lembaga lain yang dirujuk dalam skripsi ini telah disebutkan sumber kutipannya serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan atau penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan skripsi, saya bersedia menerima konsekuensi dinyatakan **TIDAK LULUS** untuk mata kuliah Skripsi yang telah saya tempuh.

Tangerang, 12 Agustus 2016

Jason Anggada

HALAMAN PERSEMBAHAN



UMMN

*Selalu berikan yang terbaik
sebagai bentuk pengabdianmu
kepada alam dan sesama manusia
juga sebagai bentuk ucapan syukurmu
kepada Tuhan Yang Maha Esa*

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, penulisan laporan Skripsi dengan judul “Implementasi Algoritma Needleman-Wunsch dan Smith-Waterman pada Penjajaran Sekuen Virus Zika” dapat diselesaikan dengan baik. Laporan ini dibuat sebagai salah satu syarat kelulusan mata kuliah Skripsi pada Program Studi Teknik Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.

Laporan Skripsi ini dapat terselesaikan karena keterlibatan dari beberapa pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Dr. Ninok Leksono, Rektor Universitas Multimedia Nusantara, yang memberikan inspirasi bagi penulis untuk mencapai sukses,
2. Kanisius Karyono, S.T., M.T., Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara, yang mengawasi dan memantau penulis untuk selalu melakukan yang terbaik,
3. Maria Irimina Prasetiyowati, S.Kom., M.T., Ketua Program Studi Teknik Informatika Universitas Multimedia Nusantara, yang selalu menerima penulis dengan baik untuk berkonsultasi,
4. Ranny, S.Kom., M.Kom., sebagai Dosen Pembimbing I, yang membimbing penelitian dan pembuatan laporan Skripsi, mengajar tata cara menulis karya ilmiah dengan benar, serta memberikan dorongan positif ketika mengalami kesulitan,
5. Eka Antonius Kurniawan, S.Kom., M.Sc., sebagai Dosen Pembimbing II, yang membimbing penelitian yang dilakukan, mengajarkan ilmu, membagikan pengalaman, serta memberikan ide dan masukan yang konstruktif kepada penulis.

6. Yustinus Widya Wiratama, S.Kom., M.Sc., yang memberikan kritik dan saran yang membangun untuk skripsi ini sebagai Dosen Penguji ketika sidang skripsi,
7. Ayah, Ibu, dan Kakak, yang senantiasa memberikan kehangatan dalam rumah tangga dan memotivasi penulis untuk selalu memberikan yang terbaik,
8. Kharis Simon, Alvin Hanjaya Tandrian, David Domarco, Samuel Christopher Santo, dan Alvin William, yang selalu menjadi sahabat yang baik dan menemani penulis dalam suka-duka kehidupan perkuliahan,
9. Julio Cristian Young dan Pandu Baraja, yang memberikan masukan yang bermanfaat dan motivasi yang positif selama pengerjaan skripsi,
10. Sebastian Limminata, Tommy Winardi, Rey Steven, Kenny Wiratama, Augy Sebastian, dan William Pratama, yang selalu memberikan dukungan kepada penulis untuk menyelesaikan skripsi meskipun terpisahkan oleh jarak, dan
11. Rekan-rekan penulis lainnya, yang tidak dapat disebutkan satu per satu, namun selalu memberikan dukungan kepada penulis dalam melakukan penelitian dan menuliskan laporan Skripsi.

Semoga laporan Skripsi ini dapat bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 12 Agustus 2016

Jason Anggada

**IMPLEMENTASI ALGORITMA
NEEDLEMAN-WUNSCH DAN SMITH-WATERMAN
PADA PENJAJARAN SEKUEN VIRUS ZIKA**

ABSTRAK

Awareness publik terhadap virus Zika telah menjadi semakin tinggi dengan adanya *outbreak* di Kepulauan Yap, Micronesia pada tahun 2007 dan di Brazil pada tahun 2015. Menurut peta persebaran virus Zika yang dirilis oleh NCBI, Indonesia telah menjadi salah satu negara yang terkena dampak virus Zika. Salah satu langkah preventif untuk menghadapi berkembangnya wabah virus Zika ini adalah dengan mengidentifikasi pola virus Zika yang muncul di Indonesia dibandingkan dengan pola virus Zika dari negara lain agar sumber virus dapat segera diketahui. Kemiripan dua buah virus dapat ditemukan dengan melakukan penjajaran sekuen. Algoritma Needleman-Wunsch digunakan untuk penjajaran global dan algoritma Smith-Waterman digunakan untuk penjajaran lokal. Dari sumber data virus Zika yang dirilis oleh NCBI, terdapat dua virus dari Indonesia, yaitu virus KU179098.1 dan virus KF258813.1. Penelitian ini menemukan bahwa virus KU179098.1 memiliki *score* terbaik dengan virus KF993678.1 dari Kanada dengan nilai 1132 bp. Virus yang sama memiliki *identity* terbaik dengan virus KX101066.1 dari Brazil dengan nilai 99.32%. Di sisi lain, virus KF258813.1 memiliki *score* terbaik dengan virus KF993678.1 dari Kanada dan virus EU545988.1 dari Micronesia dengan nilai 392 bp. Virus ini juga memiliki nilai *identity* terbaik dengan virus KX101062.1 dari Brazil, dengan nilai 99.88%.

Kata kunci: Algoritma Needleman-Wunsch, Algoritma Smith-Waterman, Penjajaran sekuen, Virus Zika.

UMMN

**IMPLEMENTATION OF NEEDLEMAN-WUNSCH
AND SMITH-WATERMAN ALGORITHM
ON ZIKA VIRUS SEQUENCE ALIGNMENT**

ABSTRACT

There has been an increase of public awareness about Zika virus with the 2007 Yap Islands Zika virus outbreak and 2015 Brazil Zika virus outbreak. According to the NCBI Zika virus heatmap, Indonesia has become one of the infected countries. One way to prevent the outbreak is to compare the pattern between Zika virus in Indonesia and Zika virus in foreign countries in order to determine the source of the infection. Similarities between two viruses can be found using sequence alignment. Needleman-Wunsch algorithm is used as the global sequence alignment and Smith-Waterman algorithm is used as the local sequence alignment. Based on the Zika virus data released by NCBI, there are two Indonesian viruses: KU179098.1 virus and KF258813.1 virus. This research has found that KU179098.1 has the highest alignment score with KKF993678 virus from Canada by 1132 bp. The same virus also has the highest alignment identity with KX101066.1 virus from Brazil by 99.32%. On the other hand, KF258813.1 virus has the highest alignment score with KF993678.1 from Canada and EU545988.1 virus from Micronesia by 392 bp. The same virus also has the highest alignment identity with KX101062.1 from Brazil by 99.88%.

Keyword: Needleman-Wunsch Algorithm, Sequence alignment, Smith-Waterman Algorithm, Zika virus.

UMMN

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI	ii
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II LANDASAN TEORI	6
2.1 Penyakit Zika	6
2.2 Nukleotida	9
2.3 Penjajaran Sekuen	10
2.4 Algoritma Needleman-Wunsch	10
2.5 Algoritma Smith-Waterman	14
2.6 National Center for Biotechnology Information (NCBI)	17
2.7 Basic Local Alignment Search Tool (BLAST)	18
2.8 Algoritma k-means clustering	19
BAB III METODOLOGI DAN PERANCANGAN SISTEM	21
3.1 Metodologi	21
3.2 Perancangan Sistem	22
BAB IV IMPLEMENTASI DAN UJI COBA	40
4.1 Spesifikasi Perangkat	40
4.2 Ekstraksi Data	41
4.3 Implementasi	43
4.4 Uji Coba	56
4.5 Analisis	59
4.6 Validasi	70
BAB V SIMPULAN DAN SARAN	72
5.1 Simpulan	72
5.2 Saran	73
DAFTAR PUSTAKA	74
DAFTAR LAMPIRAN	76

DAFTAR TABEL

Tabel 2.1 Tabel perbandingan algoritma <i>clustering</i>	20
Tabel 4.1 Tabel penjajaran virus KU179098.1 dengan <i>score</i> terbaik.....	63
Tabel 4.2 Tabel penjajaran virus KU179098.1 dengan <i>identity</i> terbaik	65
Tabel 4.3 Tabel penjajaran virus KF258813.1 dengan <i>score</i> terbaik	67
Tabel 4.4 Tabel penjajaran virus KF258813.1 dengan <i>identity</i> terbaik	69



DAFTAR GAMBAR

Gambar 1.1 Peta persebaran virus Zika	2
Gambar 2.1 Bentuk <i>virion</i> virus Zika	6
Gambar 2.2 Bentuk <i>genome</i> virus Zika	7
Gambar 2.3 <i>Guillain-Barré Syndrome</i>	8
Gambar 2.4 <i>Microcephalus</i>	8
Gambar 2.5 Susunan nukleotida dalam DNA.....	9
Gambar 2.6 <i>Similarity Matrix</i> , untuk <i>match</i> = 1 dan <i>mismatch</i> = -1	11
Gambar 2.7 <i>Pseudocode</i> Algoritma Needleman-Wunsch	12
Gambar 2.8 Tabel Needleman-Wunsch, dengan <i>traceback</i>	13
Gambar 2.9 <i>Pseudocode</i> Algoritma Smith-Waterman	15
Gambar 2.10 Tabel Smith-Waterman, dengan <i>traceback</i>	16
Gambar 2.11 Logo National Center for Biotechnology Information.....	17
Gambar 2.12 Contoh penjajaran dua virus menggunakan BLAST	18
Gambar 3.1 <i>Flowchart</i> Rangka Kerja Implementasi	23
Gambar 3.2 <i>Flowchart</i> Detil Proses Ekstraksi Data	24
Gambar 3.3 <i>Flowchart</i> Detil Proses Penyiapan Data	25
Gambar 3.4 <i>Flowchart</i> Detil Proses Memilah Virus	26
Gambar 3.5 <i>Flowchart</i> Detil Proses Penjajaran Sekuen	27
Gambar 3.6 <i>Flowchart</i> Detil Proses Penjajaran Global.....	28
Gambar 3.7 <i>Flowchart</i> Detil Proses Inialisasi Matriks	29
Gambar 3.8 <i>Flowchart</i> Detil Proses Mengisi Matriks Global	30
Gambar 3.9 <i>Flowchart</i> Detil Proses <i>Traceback</i> Global.....	31
Gambar 3.10 <i>Flowchart</i> Detil Proses Penjajaran Lokal	32
Gambar 3.11 <i>Flowchart</i> Detil Proses Mengisi Matriks Lokal.....	33
Gambar 3.12 <i>Flowchart</i> Detil Proses <i>Traceback</i> Lokal.....	35
Gambar 3.13 <i>Flowchart</i> Detil Proses Analisis dan Evaluasi	36
Gambar 3.14 <i>Flowchart</i> Detil Proses Membaca Hasil Penjajaran	37
Gambar 3.15 Desain antarmuka bagian <i>header</i>	38
Gambar 3.16 Desain antarmuka bagian kode	39
Gambar 4.1 Halaman <i>Home</i> situs NCBI.....	41
Gambar 4.2 Halaman <i>Nucleotide</i> situs NCBI.....	42
Gambar 4.3 Halaman <i>Search Results</i> situs NCBI.....	43
Gambar 4.4 Potongan kode <i>header</i>	44
Gambar 4.5 Tampilan <i>header</i> Zika virus Project.....	44
Gambar 4.6 Potongan kode <i>import libraries</i>	45
Gambar 4.7 Potongan kode <i>prepare data</i>	46
Gambar 4.8 Potongan kode dan tampilan <i>virus length scatter plot</i>	47
Gambar 4.9 Potongan kode penghitungan <i>centroid</i> dan <i>threshold</i>	48
Gambar 4.10 Informasi detil virus Zika dari Indonesia	48
Gambar 4.11 Potongan kode <i>global alignment</i>	49
Gambar 4.12 Potongan kode <i>traceback global alignment</i>	50
Gambar 4.13 Potongan kode <i>local alignment</i>	51
Gambar 4.14 Potongan kode <i>traceback local alignment</i>	53
Gambar 4.15 Potongan kode <i>execution</i>	54

Gambar 4.16 Potongan kode <i>result</i>	55
Gambar 4.17 Detil Virus KF258813.1 dan Virus KX253995.1	56
Gambar 4.18 Hasil Penjajaran Virus KF258813.1 dan Virus KX253995.1	57
Gambar 4.19 Detil Virus KU179098.1 dan Virus KX101064.1	57
Gambar 4.20 Hasil Penjajaran KU179098.1 dan Virus KX101064.1	58
Gambar 4.21 Perbandingan panjang dengan waktu untuk virus KU179098.1	59
Gambar 4.22 Perbandingan panjang dengan waktu untuk virus KF258813.1	60
Gambar 4.23 Perbandingan <i>score</i> penjajaran global dan lokal.....	61
Gambar 4.24 Perbandingan <i>score</i> penjajaran global dan lokal setelah pengurutan...	61
Gambar 4.25 <i>Scatter plot</i> untuk <i>alignment score</i> virus KU179098.1	62
Gambar 4.26 Grafik penjajaran virus KU179098.1 dengan <i>score</i> terbaik.....	64
Gambar 4.27 <i>Scatter plot</i> untuk <i>alignment identity</i> virus KU179098.1	64
Gambar 4.28 Grafik penjajaran virus KU179098.1 dengan <i>identity</i> terbaik	65
Gambar 4.29 <i>Scatter plot</i> untuk <i>alignment score</i> virus KF258813.1	66
Gambar 4.30 Grafik sepuluh penjajaran virus KF258813.1 dengan <i>score</i> terbaik....	67
Gambar 4.31 <i>Scatter plot</i> untuk <i>alignment identity</i> virus KF258813.1	68
Gambar 4.32 Grafik penjajaran virus KF258813.1 dengan <i>identity</i> terbaik.....	69
Gambar 4.33 Grafik perbandingan validasi BLAST untuk virus KU179098.1.....	70
Gambar 4.34 Grafik perbandingan validasi BLAST untuk virus KF258813.1	71

