

ABSTRAK

Ilmu Bioinformatika meneliti tentang perubahan yang dialami oleh DNA, serta membantu memberikan tanda terhadap mutasi genetika yang terjadi. Untuk membandingkan sekuens DNA dan mencari tahu bagaimana kedua DNA memiliki kesamaan dapat digunakan algoritma-algoritma yang dapat mengolah pensejajaran DNA baik secara global maupun secara lokal. Pensejajaran secara global dilakukan dengan membandingkan semua karakter di dalam sekuens, sementara pensejajaran lokal hanya mengambil potongan dari sekuens dan membandingkannya. Pensejajaran global menjadi dasar penelitian laporan ini dengan menggunakan dua algoritma yaitu Needleman-Wunsch dan Lempel-Ziv. Kedua algoritma ini bekerja dengan membangun matriks skor dan mensejajarkan sekuens dari hasil matriks yang dibuat. Pengujian dilakukan dengan mengambil secara acak sekuens DNA dengan panjang kurang dari 1000 karakter hingga panjang melebihi 1000 karakter. Algoritma Needleman-Wunsch unggul dengan kecepatan proses hingga 1 *miliseconds* untuk *dataset* kurang dari 1000 karakter dan 42 *miliseconds* untuk *dataset* lebih dari 1000 karakter. Meskipun algoritma Lempel-Ziv memerlukan waktu untuk pembentukan frase, kecepatan algoritma Lempel-Ziv rupanya lebih cepat daripada algoritma Needleman-Wunsch untuk kasus sekuens DNA yang memiliki frase sempurna.

Kata kunci: DNA, bioinformatika, sekuens, Needleman-Wunsch, Lempel-Ziv, algoritma pensejajaran DNA, frase sempurna

ABSTRACT

Bioinformatics research is currently working on the changing of the DNA information, and marking the mutation for the DNA. For comparing DNA and finding out how two DNA can have similarities, bioinformatics using algorithms that works in global alignment and local alignment. The global alignment is comparing all the characters in sequence while the local only take a piece of characters from the alignment. This study proposes two algorithms for processing the DNA sequence in global alignment. The algorithms are Needleman-Wunsch and Lempel-Ziv algorithms. These algorithms work with building a scoring matrix and create an alignment based on the matrix. The experiment is conducted by testing DNA sequences randomly with the length less than 1000 characters and more than 1000 characters. Needleman-Wunsch leading with processing speed up to 1 miliseconds for less than 1000 character dataset and 42 miliseconds for more than 1000 characters dataset, while Lempel-Ziv is leading the processing speed on specific case of perfect phrase in DNA sequence.

Keywords: DNA, Bioinformatics, sequence, Needleman-Wunsch, Lempel-Ziv, sequence alignment algorithms, perfect phrase.



DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS LAPORAN PENELITIAN.....	ii
PERNYATAAN PUBLIKASI LAPORAN PENELITIAN	iii
PRAKATA.....	iv
ABSTRAK	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xiii
DAFTAR NOTASI/ LAMBANG.....	xiv
DAFTAR SINGKATAN	xvi
DAFTAR ISTILAH	xvii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Pembahasan	3
1.4 Ruang Lingkup.....	3
1.5 Sumber Data.....	4
1.6 Sistematika Penyajian	4
BAB 2 KAJIAN TEORI	6
2.1 Bioinformatika	6
2.2 DNA, RNA, dan Protein	6
2.3 Rekayasa Perangkat Lunak	7
2.4 Algoritma Needleman-Wunsch.....	8

2.5 Algoritma Lempel-Ziv	9
2.6 Skor Dan Kecepatan Algoritma	10
BAB 3 ANALISIS DAN RANCANGAN SISTEM.....	11
3.1 Analisis Program.....	11
3.2 Needleman-Wunsch	11
3.2.1 Scoring Matrix	11
3.2.2 Traceback	14
3.2.3 Alignment.....	14
3.3 Lempel-Ziv.....	15
3.3.1 Lempel-Ziv Factorisation Phrase	15
3.4 Substitution Matrix	18
3.4.1 Gambaran Aplikasi.....	18
3.5 Gambaran Keseluruhan.....	19
3.5.1 Persyaratan Antarmuka Eksternal	19
3.5.2 Persyaratan Antarmuka Perangkat Lunak	19
3.5.3 Persyaratan Antaruka Perangkat Keras	20
3.6 Pemodelan Perangkat Lunak.....	20
3.6.1 Use case Diagram.....	20
3.6.2 Package Dan Class Diagram	21
3.6.3 Skenario	22
3.6.3.1 Skenario Sistem: Input Sekuens	22
3.6.3.2 Skenario Sistem: Pilih Metode Pensejajaran	23
3.6.3.3 Skenario Sistem : Menentukan Nilai Pensejajaran	25
3.6.4 Activity Diagram.....	25
3.6.4.1 Activity Diagram: Input Sekuens	26
3.6.4.2 Activity Diagram: Pilih Metode Pensejajaran.....	26

3.7 Rencana Pengujian	27
BAB 4 IMPLEMENTASI.....	28
4.1 Hasil Tampilan	28
4.2 Tampilan Awal Aplikasi	28
4.3 Tampilan Progress Dan Output.....	29
4.4 Implementasi Aplikasi	30
4.4.1 Scoring Matrix	30
4.4.2 Factor Sequence	34
4.4.3 Needleman-Wunsch	36
4.4.3.1 Compute Matrix	36
4.4.3.2 Pairwise Alignment Needleman-Wunsch	38
4.4.4 Lempel-Ziv.....	40
BAB 5 PENGUJIAN	42
5.1 Hasil Pengujian	42
5.2 Pengujian Sekuens Maksimal 1000 Karakter	42
5.3 Pengujian Sekuens Di Atas 1000 Karakter	47
5.4 Pengujian Spesifik Lempel-Ziv Factorisation Phrase Terhadap Nilai Gap	51
BAB 6 SIMPULAN DAN SARAN.....	53
6.1 Simpulan	53
6.2 Saran.....	54
DAFTAR PUSTAKA	55
RIWAYAT HIDUP PENULIS	57

DAFTAR GAMBAR

Gambar 2.1 Dogma sentral biologi molekuler	7
Gambar 3.1 Proses Awal Matriks Skor.....	12
Gambar 3.2 Proses Pengisian Tabel Matriks Skor.....	13
Gambar 3.3 Hasil Pengisian Matriks Skor.....	14
Gambar 3.4 Hasil Alignment Needleman-Wunsch.....	15
Gambar 3.5 Proses Pengisian Matriks Skor Lempel-Ziv.....	17
Gambar 3.6 Hasil Alignment Lempel-Ziv	17
Gambar 3.7 Contoh Substitution Matrix BLOSUM	18
Gambar 3.8 Use case Sistem Bioinformatika	20
Gambar 3.9 Package Bioinformatika	21
Gambar 3.10 Class Diagram Bioinformatika.....	22
Gambar 3.11 Input Sekuens	26
Gambar 3.12 Pilih metode pensejajaran.....	26
Gambar 3.13 Menentukan nilai pensejajaran.....	27
Gambar 4.1 Tampilan Awal Aplikasi Bioinformatika.....	29
Gambar 4.2 Tampilan Progress.....	30
Gambar 4.3 Tampilan Output	30
Gambar 4.4 Kode Program Untuk Proses Scoring Matrix.....	33
Gambar 4.5 Contoh Proses Scoring Matrix	33
Gambar 4.6 Kode Program Untuk Factor Sequence.....	35
Gambar 4.7 Contoh Factor Sequence.....	35
Gambar 4.8 Kode Program Untuk Compute Matrix.....	37
Gambar 4.9 Contoh Proses Compute Matrix	37
Gambar 4.10 Kode Program Untuk Pairwise Alignment Needleman-Wunsch....	39
Gambar 4.11 Contoh Proses Sequence Alignment Needleman-Wunsch.....	39
Gambar 4.12 Kode Program Untuk Pairwise Alignment Lempel-Ziv	41
Gambar 4.13 Contoh Pairwise Sequence Lempel-Ziv	41
Gambar 5.1 Grafik Waktu Untuk Pengujian Sekuens Maksimal 1000 Karakter .	44
Gambar 5.2 Grafik Panjang Sekuens Untuk Tiap Test Case	45
Gambar 5.3 Grafik Skor Alignment Terhadap Panjang Sekuens	45

Gambar 5.4 Grafik Waktu Proses Terhadap Panjang Sekuens	46
Gambar 5.5 Grafik Waktu Proses Terhadap Test Case.....	49
Gambar 5.6 Grafik Panjang Sekuens Terhadap Test Case	50
Gambar 5.7 Grafik Skor Terhadap Panjang Sekuens.....	50
Gambar 5.8 Grafik Waktu Proses Terhadap Panjang Sekuens	51
Gambar 5.9 Hasil Lempel-Ziv Alignment Sekuens Hepatitis A dan Hepatitis B.	52

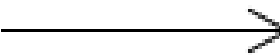


DAFTAR TABEL

Tabel 2.1 Jenis perangkat lunak dan contohnya.....	7
Tabel 3.1 Lempel-Ziv Factorisation Phrase	16
Tabel 3.2 Skenario Input Sekuens.....	23
Tabel 3.3 Skenario Pilih Metode Pensejajaran	24
Tabel 3.4 Skenario Menentukan Nilai Pensejajaran	25
Tabel 5.1 Pengujian Sekuens Maksimal 1000 Karakter	42
Tabel 5.2 Pengujian Sekuens Di Atas 1000 Karakter.....	47
Tabel 5.3 Contoh Faktorisasi Sempurna Lempel-Ziv	48



DAFTAR NOTASI/ LAMBANG

Jenis	Notasi/ Lambang	Nama	Arti
UML		<i>Association</i>	Relasi antar kelas dengan makna umum, asosiasi biasanya juga disertai dengan <i>multiplicity</i>
UML		<i>Directed association</i>	Relasi antar kelas dengan makna kelas yang satu digunakan oleh kelas yang lain, asosiasi biasanya juga disertai dengan <i>multiplicity</i>
UML		<i>Use case</i>	Fungsionalitas yang disediakan sistem sebagai unit-unit yang saling bertukar pesan antar unit atau aktor.
UML		<i>Actor</i>	Orang, proses, atau sistem lain yang berinteraksi dengan sistem informasi yang akan dibuat di luar sistem informasi itu sendiri, jadi walaupun gambar aktor adalah gambar orang belum tentu aktor adalah orang
UML		Status awal	Status awal aktivitas sistem, sebuah diagram aktivitas memiliki sebuah status awal
UML		Status Akhir	Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki status akhir

Jenis	Notasi/ Lambang	Nama	Arti
UML		Aktivitas	Aktivitas yang dilakukan sistem, biasanya diawali dengan kata kerja
UML		Percabangan	Asosiasi percabangan di mana jika ada pilihan aktivitas lebih dari satu

Referensi:

Notasi/ Lambang ERD dan DFD dari buku Rekayasa Perangkat Lunak [1]



DAFTAR SINGKATAN

DNA	<i>Deoxyribonucleic Acid</i>
RNA	<i>Ribonucleic Acid</i>
mRNA	<i>Messenger Ribonucleic Acid</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
LASTZ	<i>Local Alignment Search Tool - Z</i>
YASS	<i>Yet Another Similarity Search</i>
RPL	Rekayasa Perangkat Lunak
NCBI	National Center for Biotechnology Information



DAFTAR ISTILAH

<i>Sequence</i>	Sekuens atau kumpulan karakter DNA yang berupa urutan huruf-huruf.
<i>Matrix</i>	Matriks, implementasi penyusunan karakter-karakter menjadi kotak yang tersusun sehingga dapat dihitung pensejajarannya.
<i>Alignment</i>	Pensejajaran, teknik mencari kesejajaran dari sekuens DNA.
<i>Phrase</i>	Frase
<i>Data Compression</i>	Kompresi data menjadi lebih singkat/kecil.
<i>Perfect Factorisation</i>	Faktorisasi sempurna untuk menandakan penggunaan dalam algoritma Lempel-Ziv

