

Perkembangan Bioinformatics dalam Ruang Lingkup Ilmu Komputer

Irma Yunita, Kristian Tjandradiredja, Seng Hansun

Program Studi Teknik Informatika, Universitas Multimedia Nusantara, Tangerang, Indonesia
irma.yunita@student.umn.ac.id, kristian.tjandradiredja@student.umn.ac.id

Diterima 19 Mei 2016

Disetujui 10 Juni 2016

Abstract—Bioinformatics is a research topic which is very attractive in this era. Bioinformatics combines biology and computer science. It means that biology problem can be solved by using methods and techniques in solving problems in computer science. There are some terms of bioinformatics such as DNA, RNA, and protein. In addition, there are several methods described in this paper, which can be used in solving the problem bioinformatics following with the tools and software which are available and free in internet.

Keywords—bioinformatics, tools of bioinformatics, methods of bioinformatics

I. PENDAHULUAN

Pada era teknologi yang semakin berkembang ini, kebutuhan akan informasi semakin dibutuhkan oleh masyarakat secara luas. Selain itu, masalah yang berhubungan dengan dunia biologis atau lebih dikenal dengan *bioinformatics*, menjadi topik yang diminati pada masa kini [1].

Bioinformatics diciptakan oleh Paulen Hogeweg pada tahun 1979 sebagai ilmu yang mempelajari tentang proses informasi pada sistem biotik [2]. Menurut [2], *bioinformatics* dapat didefinisikan sebagai aplikasi dari teknologi komputer untuk mengelola informasi biologis.

Berikut ini alasan *bioinformatics* menjadi tema sangat menarik untuk dibahas [3].

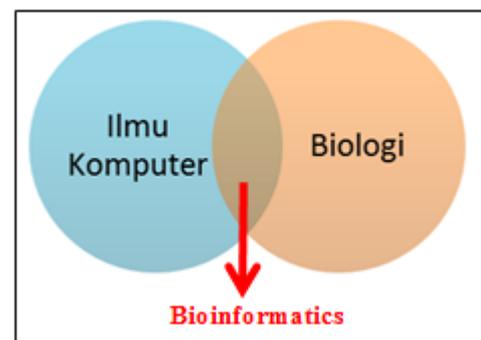
- Biologi merupakan masalah yang sangat penting karena berhubungan dengan manusia.
- Menggunakan data dengan jumlah yang banyak.
- Bertujuan untuk menyelesaikan masalah (solusi).
- Jika berhasil menyampaikan solusi, akan memperoleh penghargaan.

Bioinformatics memiliki tujuan utama untuk meningkatkan pengertian tentang proses secara biologis [2]. Selain itu, *bioinformatics* juga dapat digunakan untuk menyelesaikan masalah yang berkaitan dengan manusia. Oleh karena itu, di sini dibahas lebih rinci tentang *bioinformatics* beserta dengan algoritma-algoritma yang dapat digunakan dalam menyelesaikan

masalah biologis dan penerapannya dengan pendekatan pada ruang lingkup ilmu komputer.

II. DEFINISI BIOINFORMATICS

Bioinformatics merupakan pembuatan konsep biologis dengan istilah molekul (memiliki makna yang sama dengan kimia fisik) dan menggunakan teknik informasi (diturunkan berdasarkan disiplin seperti menggunakan matematika, ilmu komputer dan statistik) untuk memahami dan mengorganisasi informasi terkait dengan molekul tersebut pada skala yang luas [4]. Berdasarkan Gambar 1, secara lebih sederhana, *bioinformatics* merupakan irisan antara ilmu komputer dan biologi.



Gambar 1. Definisi Bioinformatics

III. TERMINOLOGI BIOINFORMATICS

Berikut ini dijelaskan beberapa terminologi dari *bioinformatics* menurut [5].

A. Gen atau Genome

Gen membawa informasi turunan dari sebuah organisme yang direpresentasikan dalam setiap sel. Gen direpresentasikan dalam bentuk molekul DNA. Setiap molekul DNA merupakan rantai panjang yang terdiri dari struktur asam nukleat yang terdiri dari empat tipe, yaitu *Adenine*, *Cytosine*, *Guanine*, *Thymine*. Keempat tipe tersebut biasa disingkat menjadi A, C, G, T.

Informasi yang diperoleh dari gen dapat saling dibandingkan dengan gen lainnya dari spesies berbeda.

Hal ini dapat menjelaskan proses dan mekanisme dari evolusi.

B. Genetik

Genetik merupakan studi atau ilmu yang mempelajari tentang turunan biologis dari suatu organisme. Hal utama dari genetik ini merupakan konsep dari gen karena gen yang dapat diwariskan.

C. Sel

Sel merupakan unit terkecil penyusun organisme. Manusia terdiri dari triliun sel yang dapat melakukan mitosis dan meiosis, serta membawa informasi tentang sel yang dapat menghasilkan keturunan.

D. Protein

Protein merupakan molekul biologis yang sangat besar yang tersusun dari untaian molekul-molekul yang lebih kecil, biasa disebut asam amino (*amino acids*). Protein dapat membuat struktur sel, seperti rambut, kulit, dan lain-lain.

E. DNA

DNA atau *Deoxyribonucleic Acid* merupakan untaian kode genetik yang terdiri dari {A, C, G, T} yang membawa informasi keturunan dari gen sebelumnya.

F. RNA

RNA atau *Ribonucleic Acid* merupakan untaian ganda dan pendek kode genetik yang terdiri dari *Adenine, Cytosine, Guanine, Urasil* ({A, C, G, U}) yang membawa informasi keturunan dari gen sebelumnya dan berperan penting dalam pembentukan protein.

IV. ANALOGI BIOLOGI DAN ILMU KOMPUTER

Analogi biologi dan ilmu komputer ini dapat diajukan untuk mempermudah penalaran tentang *bioinformatics* terutama dalam hal pemrosesan gen dan protein dalam komputer. Menurut [1], analogi antara ilmu biologi dan ilmu komputer diuraikan sebagai berikut.

- Gen G pada komponen DNA berperan penting pembentukan protein P.
- Suatu *Interpreter* I mampu memproses setiap gen yang mungkin digunakan oleh protein P sebagai satu dari komponen *interpreter* itu sendiri.
- Hal tersebut memberi dampak jika protein P tidak diproduksi ke dalam mesin dari I, tidak ada proses interpretasi yang dapat berlangsung.
- Situasi di atas dapat disimulasikan pada komputer dengan memanfaatkan proses yang berjalan bersamaan, yaitu terdiri dari banyak interupsi untuk mengendalikan eksekusi program.
- Implementasi tersebut menggunakan interpreter dimana pengujian pertama-tama dilakukan pada semua komponen yang telah diproduksi. Pengujian

hasil tersebut hanya terjadi jika proses pengujian terhadap semua komponen telah dilaksanakan dan interupsi berlangsung hingga proses produksi selesai. Eksekusi bagian program dilaksanakan setelah seluruh komponen diproduksi dan interupsi terjadi sebelum hasil produksi selesai.

V. PENELITIAN TERKAIT BIOINFORMATICS

Sejak penemuan pertama pada 1980-an, penelitian terkait *bioinformatics* semakin berkembang [7]. Penggabungan antara permasalahan dalam ilmu biologi dan ilmu komputer, memiliki daya tarik yang kuat dalam era perkembangan teknologi informasi yang semakin pesat pada masa kini. Penelitian terkait *bioinformatics* menghubungkan antara permasalahan dalam ilmu biologi yang diselesaikan secara komputerisasi. Berikut ini disajikan beberapa penelitian sebelumnya terkait dengan *bioinformatics*.

A. Bioinformatics dalam Penelitian Penyakit Kanker

Pengelompokan atau penentuan jenis penyakit kanker tertentu merupakan salah satu permasalahan *bioinformatics*. Berdasarkan penelitian yang telah dilakukan oleh [6], dalam kasus pengelompokan kanker tertentu yang berdasarkan pola dari ekspresi gen, dapat bekerja dengan baik untuk meningkatkan fungsionalitas analisis terhadap ciri-ciri kanker tertentu.

Penelitian tersebut [6] menggunakan *microarray data* dan protein basis data. Jika fungsi gen tidak dapat diperoleh dari basis data secara terbuka, dapat digunakan metode pencarian berbasis homologi, yang mencakup konteks *string*, PFP, dan PSORT [6]. Perangkat *bioinformatics* dan basis data yang digunakan dalam penelitian tersebut di atas tersedia dalam halaman *website* <http://dragon.bio.purdue.edu/bioinfolinks>.

B. OrthologID

Penelitian selanjutnya dikembangkan oleh [7] yang menggunakan tanaman sebagai objek penelitian. Berdasarkan [7], OrthologID merupakan aplikasi berbasis web yang mengotomatisasi prosedur besar dari ortologi gen dan karakteristik dari kerangka filogenetik. Dengan demikian, dapat membuat kepastian ortologi yang berdasarkan karakteristik pada skala gen menjadi memungkinkan.

C. RNA Sampler

Menurut [8], tidak adanya penulisan kode pada RNA dan aturan RNA secara terstruktur menjadi pendorong terhadap pentingnya peran dalam aturan gen dan fungsi seluler lainnya. Berdasarkan penelitian yang telah dilakukan oleh [8], dapat memprediksi struktur RNA sekunder yang secara umum.

Algoritma yang disajikan oleh [8], dapat menemukan struktur umum di antara dua kejadian dari percobaan

probabilitas yang diperoleh dari hasil perhitungan konservasi dengan membandingkan antara dua kejadian (perhitungan berdasarkan probabilitasnya). Algoritma yang diajukan [8] memiliki kelebihan yaitu mampu memprediksi struktur metode dengan tingkat kerentanan dan ketegasan dengan cepat dan mampu memberikan alasannya.

RNA *Sampler* tersedia dalam <http://ural.wustl.edu/software.html> dan penelitian tersebut dibangun dengan menggunakan bahasa pemrograman C. Berdasarkan percobaan yang telah dilakukan oleh [8], dengan menggunakan RNA *Sampler*, mampu memprediksi kebutuhan struktur sekunder dari RNA secara umum.

VI. METODE YANG SERING DIGUNAKAN DALAM BIOINFORMATICS

Berikut ini dijelaskan beberapa metode yang sering digunakan dalam masalah *bioinformatics* menurut [1].

A. Membandingkan Urutan

Diberikan data yang tersedia berjumlah sangat banyak, terdapat kebutuhan penting untuk mengembangkan algoritma yang sesuai dalam membandingkan urutan data yang sangat besar. Algoritma ini mengizinkan adanya penghapusan (*delete*), penambahan (*insertion*), dan penggantian (*replacements*) dari simbol-simbol yang merepresentasikan inti dari DNA (nukleotida) atau asam amino (protein) untuk adanya perubahan secara alami.

B. Membangun Evolusi (Phylogenetic Trees)

Struktur pohon (*tree*) ini biasanya dibuat setelah proses membandingkan urutan untuk setiap jenis organisme selesai. Pengelompokan *tree* berdasarkan tingkat dari kemiripannya. Pengelompokan tersebut dianggap sebagai panduan untuk penalaran tentang bagaimana urutan tersebut telah diubah melalui proses evolusi. Misalnya, pengelompokan homologi dari kesamaan tertentu, dan mungkin mengesampingkan asumsi kesalahan (*error*) yang bertentangan dalam proses evolusi.

C. Mendeteksi Pola dalam Urutan

Ada bagian-bagian tertentu dari urutan DNA dan asam amino yang perlu dideteksi. Dua contoh utama yaitu gen dalam DNA dan menentukan komponen bagian dari urutan asam amino (struktur sekunder). Terdapat beberapa cara untuk melakukan proses tersebut. Biasanya proses tersebut dilakukan berdasarkan *machine learning* berikut dengan probabilitas dari aturan bahasa (*grammar*) dan jaringan saraf (*neural networks*).

D. Menentukan Struktur 3D dari Urutan

Masalah dalam *bioinformatics* yang berhubungan dengan urutan ke struktur 3D merupakan masalah komputasi yang sulit. Penentuan bentuk RNA dari urutan tersebut, membutuhkan algoritma dengan

tingkat kompleksitas ($O(n)$) pangkat tiga.

Kesimpulan dari bentuk protein yang diproduksi dari urutan asam amino merupakan masalah yang hingga saat ini belum dapat dipecahkan.

E. Menyimpulkan Aturan Sel

Fungsi gen atau protein terbaik digambarkan oleh perannya dalam metabolisme atau jalur sinyalnya. Gen berinteraksi satu sama lain dan protein juga dapat mencegah atau membantu dalam produksi protein lainnya. Model perkiraan yang tersedia untuk aturan sel dapat berupa diskrit atau kontinu. Pada umumnya, aturan tersebut membedakan antara simulasi sel dan pemodelannya. Jumlah terakhir data digunakan untuk menyimpulkan data yang sudah tidak dipakai dari data eksperimen (biasa disebut sebagai *microarray*). Proses ini biasanya disebut *reverse engineering*.

F. Menentukan Fungsi Protein dan Persiapan Metabolisme

Hal ini adalah salah satu dari daerah yang paling menantang dari *bioinformatics* dan yang tidak ada data yang cukup tersedia untuk menyelesaikannya. Tujuannya adalah untuk menafsirkan penjelasan manusia tentang fungsi protein dan juga untuk mengembangkan *database* yang mewakili grafik yang dapat dilihat untuk keberadaan *node* (menentukan reaksi) dan jalur (menentukan urutan reaksi).

G. Menyusun Potongan DNA (DNA Fragments)

Potongan DNA yang disediakan oleh mesin yang dapat mengurutkan, dirakit menggunakan komputer. Hal tersulit dari kumpulan tersebut adalah DNA memiliki banyak daerah yang berulang-ulang dan potongan yang sama mungkin dimiliki oleh daerah yang berbeda. Algoritma untuk perakitan DNA yang banyak digunakan oleh perusahaan-perusahaan besar seperti *Former Celera*.

H. Menggunakan Bahasa Scripting

Banyak aplikasi di atas sudah tersedia pada *website-website* secara umum. Penggunaannya membutuhkan *scripting* yang menyediakan data untuk aplikasi, menerimanya kembali, dan kemudian menganalisisnya.

Yang membedakan masalah *bioinformatik* dari orang lain adalah ukuran besar data dan kualitasnya. Hal tersebut menunjukkan bahwa kebutuhan untuk perkiraan solusi yang tepat.

Perlu diperhatikan bahwa beberapa masalah dalam *bioinformatics* merupakan masalah optimasi (*optimization problem*). Solusi untuk masalah tersebut biasanya komputasi mahal. Salah satu metode yang efisien yang dikenal dalam masalah optimasi yaitu pemrograman dinamis (*dynamic programming*). Hal tersebut menjelaskan alasan teknik ini sering digunakan dalam *bioinformatics*. Pendekatan lain seperti *branch-and-bound* juga digunakan dalam mengatasi masalah optimasi, tetapi mereka dikenal memiliki kompleksitas yang lebih tinggi dari *dynamic programming*.

VII. EVALUASI TERHADAP PERANGKAT SIMULASI BIOINFORMATICS

Terdapat beberapa faktor yang harus diperhatikan dalam menentukan perangkat (*tools*) simulasi yang baik untuk digunakan. Faktor-faktor tersebut yaitu penerapan (*applicability*), kegunaan (*usability*), kelebihan dan kekurangan, pendaftaran dan ketersediaan buku panduan, dan estimasi parameter [9]. Faktor-faktor tersebut diuji dalam *biochemical network*.

Aplikasi yang digunakan sebagai obyek evaluasi, yaitu GENESIS/Kinetikit, Jarnac/JDesigner, dan Gepasi. Hasil evaluasi berdasarkan faktor-faktor di atas dijelaskan lebih rinci pada Tabel 1 [9].

Berikut ini dijelaskan lebih lanjut tentang faktor-faktor yang digunakan sebagai parameter dalam mengukur aplikasi yang baik untuk digunakan dalam simulasi pada masalah *bioinformatics* khususnya *biochemical network*.

A. Penerapan (Applicability)

Penerapan ini menjelaskan daerah aplikasi utama dari simulator. Semua simulator mendukung untuk melakukan simulasi berdasarkan waktu.

B. Kegunaan (Usability)

Kegunaan mendeskripsikan kemudahan dalam menggunakan aplikasi untuk dipelajari dan masalah yang muncul ketika melakukan pengujian. *Graphical User Interface* (GUI) sangat dibutuhkan untuk memudahkan penggunaan dalam melakukan simulasi.

C. Kelebihan dan Kekurangan

Hal ini menunjukkan hasil evaluasi terhadap program aplikasi simulasi. Selain itu, dipertimbangkan pula keandalan dan kompatibilitas sebagai nilai yang sangat penting dalam sebuah aplikasi simulasi.

D. Pendaftaran dan Ketersediaan Buku Panduan

Sebagian besar dari aplikasi simulasi tersedia secara gratis, tetapi lebih dari 20 *tools* dibutuhkan registrasi sebagai pelanggan. Pada beberapa kasus, registrasi telah dilakukan dengan menerapkan hak penggunaan.

E. Estimasi Parameter (Parameter Estimation)

Sesuai dengan kemungkinan estimasi komputasi nilai parameter model (model fitting).

VIII. TOOLS PENTING YANG DAPAT DIGUNAKAN DALAM BIOINFORMATICS

Berikut ini dijelaskan beberapa *tools* penting yang sering digunakan dalam masalah *bioinformatics* beserta dengan referensi untuk mendapatkan *tools* tersebut [2] dalam Tabel 2.

IX. SIMPULAN

Bioinformatics menjadi topik yang sangat diminati pada masa kini. *Bioinformatics* menghubungkan antara ilmu biologi dan ilmu komputer seperti diilustrasikan pada Gambar 1.

Sejak penemuan pertamanya, *bioinformatics* terus mengalami perkembangan yang sangat signifikan. Penelitian terkait dalam *bioinformatics* disajikan pada Bab V. Disajikan pula metode yang digunakan beserta dengan hasil yang diperoleh dari penelitian-penelitian sebelumnya.

Terdapat beberapa terminologi dalam ilmu biologi yang digunakan dalam mengolah data pada komputer seperti DNA, RNA, dan protein. Metode pengolahan data dalam *bioinformatics* disajikan pada Bab VI.

Penelitian terhadap *bioinformatics* hingga saat ini masih dalam tahap perkembangan. Dalam memudahkan implementasi *bioinformatics*, terdapat *tools* dan aplikasi simulasi yang tersedia secara gratis.

Tabel 2. Beberapa *tools* dalam *bioinformatics*

Daerah Penelitian Bioinformatics	Tool (Aplikasi)	Referensi
Sequence Alignment	BLAST	http://blast.ncbi.nlm.nih.gov/Blast.cgi
	CS-BLAST	ftp://toolkit.lmb.uni-muenchen.de/csblast/
	HMMER	http://hmm.janelia.org/
	FASTA	www.ebi.ac.uk/fasta33
Multiple Sequence Alignment	MSAProbs	http://msaprobs.sourceforge.net/
	DNA Alignment	http://www.fluxus-engineering.com/align.htm
	MultAlin	http://multalin.toulouse.inra.fr/multalin/multalin.html
	DiAlign	http://bibiserv.techfak.uni-bielefeld.de/dialign/
Menemukan Gen	GenScan	genes.mit.edu/GENSCAN.html
	GenomeScan	http://genes.mit.edu/genomescan.html
	GeneMark	http://exon.biology.gatech.edu/
Analisis Protein Domain	Pfam	http://pfam.sanger.ac.uk/
	BLOCKS	http://blocks.jherc.org/
	ProDom	http://prodom.prabi.fr/prodom/current/html/home.php

Daerah Penelitian Bioinformatics	Tool (Aplikasi)	Referensi
Identifikasi Pola	Gibbs Sampler	http://bayesweb.wadsworth.org/gibbs/gibbs.html
	AlignACE	http://atlas.med.harvard.edu/
	MEME	http://meme.sdsc.edu/

Daerah Penelitian Bioinformatics	Tool (Aplikasi)	Referensi
Analisis Gen	SLAM	http://bio.math.berkeley.edu/slam/
	Multiz	http://www.bx.psu.edu/miller_lab/
Menemukan Motif	MEME/MAST	http://meme.sdsc.edu
	eMOTIF	http://motif.stanford.edu

LAMPIRAN

Tabel 2. Evaluasi terhadap aplikasi simulasi dalam *bioinformatics*

Faktor	GENESIS/Kinekit	Jarnac/JDesigner	Gepasi
Applicability	Merancang dan melakukan simulasi pada <i>biochemical networks</i> dan <i>neuronal systems</i>	Merancang melakukan simulasi dalam <i>biochemical networks</i>	Simulasi untuk <i>biochemical networks</i> , estimasi dari nilai model parameter, dan <i>metabolic control analysis</i>
Usability	Membutuhkan pengalaman dalam menggunakan GUI dan <i>script language</i>	JDesigner merupakan <i>intuitive</i> GUI dalam merancang jaringan dan mengendalikan Jarnac	<i>Package</i> aplikasi yang mudah digunakan
Kelebihan	Baik digunakan dalam hal komputasi dan efisien	Mendukung <i>external</i> SBW <i>interface</i> , SBML <i>import</i> dan <i>export</i>	Baik digunakan dalam hal komputasi dan efisien, tersedia metode yang beragam, SBML <i>import</i> dan <i>export</i>
Kekurangan	GUI tidak mudah digunakan oleh kaum awam atau pemula	Tidak dapat memanfaatkan selain untuk <i>linear external stimulus</i>	Tidak ada <i>tool</i> perancangan
Pendaftaran dan Ketersediaan Buku Panduan	Tidak membutuhkan pendaftaran dan tersedia buku panduan	Gratis dan tersedia buku panduannya	Tidak dibutuhkan pendaftaran dan buku panduan tersedia
Estimasi Parameter	Sebanyak 3 metode tersedia	Tidak ada metode yang tersedia	Sebanyak 12 metode tersedia

DAFTAR PUSTAKA

- [1] C. Jacques, "Bioinformatics - An Introduction For Computer Scientists," ACM Computing Surveys, vol. XXXVI, No. 2, Juni 2004, hal. 122-158.
- [2] R. Khalid, "Application Of Data Mining In Bioinformatics," Indian Journal of Computer Science and Engineering, vol. I, No. 2, hal. 114-118.
- [3] L. Dan, "Introduction Of Bioinformatics"[online], tersedia dalam https://www.lehigh.edu/~inbios21/PDF/Fall2010/Lopresti_10082010.pdf, diakses 17 Mei 2016.
- [4] N.M. Luscombe, D. Greenbaum, M. Gerstein, "What is bioinformatics? An introduction and overview," dalam Yearbook of Medical Informatics 2001, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, USA.
- [5] M. T. Sabu, "Bioinformatics"[online], tersedia dalam <https://arxiv.org/ftp/arxiv/papers/0911/0911.4230.pdf>, diakses 17 Mei 2016.
- [6] K. Daisuke, D.Y. Yifeng, H. Troy, "Bioinformatics Resources For Cancer Research With An Emphasis On Gene Function And Structure Prediction Tools," Cancer Informatics, 2006, hal. 25-35.
- [7] C. C. Joanna, dkk, "OrthologID: Automation Of Genome-Scale Ortholog Identification Within A Parsimony Framework," Oxford University Press, vol. 22, No. 6, Januari 2006, hal. 699-707.
- [8] X. Xing, J. Yongmei, D. S. Gary, "RNA Sampler: A New Sampling Based Algorithm For Common RNA Secondary Structure Prediction And Structural Alignment," Oxford University, vol. 23, No. 15, Mei 2007, hal. 1883-1891.
- [9] P. Antti dkk, "Simulation Tools For Biochemical networks: Evaluation Of Performance and Usability," Bioninformatics Advance Access, 9 September 2004, Oxford University, hal. 1-8.