

Menemukan Kandidat Obat untuk Penyakit Alzheimer's dengan Approximate String Matching pada PharmGKB Database

Agung Baptiso Sorlawan - 13513043
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
agungbsorlawan@students.itb.ac.id

Abstract—Alzheimer Disease (AD) bukan penyakit menular, melainkan merupakan sejenis sindrom dengan apoptosis sel-sel otak pada saat yang hampir bersamaan, sehingga otak tampak mengerut dan mengecil. Alzheimer biasanya diikuti dengan demensia atau penurunan fungsional yang disebabkan oleh kelainan yang terjadi pada otak. Meskipun beberapa antidemensia seperti donepezil, rivastigmine, galantamine, and memantine telah dikembangkan, tetapi tingkat keefektifannya masih jauh dari harapan. Saat ini, the Single Nucleotide Polymorphisms (SNPs) menjadi acuan untuk pembuatan obat untuk masing-masing penderita AD. Banyak database kefarmasian yang menyediakan informasi yang mengasosiasikan SNPs dengan respon obat, indikasi, dan gen yang harus diterapi. Pada makalah ini, akan dijelaskan penggunaan string matching untuk pencarian dalam database farmasi genomik (*pharmacogenomics*). Hal ini mungkin dapat meningkatkan bidang farmasi genomik dalam menentukan obat untuk masing-masing penderita.

Index Terms—Alzheimer's, AD, DNA, SNPs, PharmGKB, Approximate String Matching.

I. INTRODUCTION

Alzheimer's disease (AD), bentuk terbanyak dari demensia pertama kali ditemukan pada tahun 1906. Pada 2006, terdapat sekitar 26,6 juta pasien AD di dunia^[1]. Meskipun AD telah diidentifikasi sejak 1 abad lalu, namun tidak ada cara mencegahnya. Oleh karena itu, pencarian obat untuk AD menjadi sebuah tantangan tantangan di masa kini.

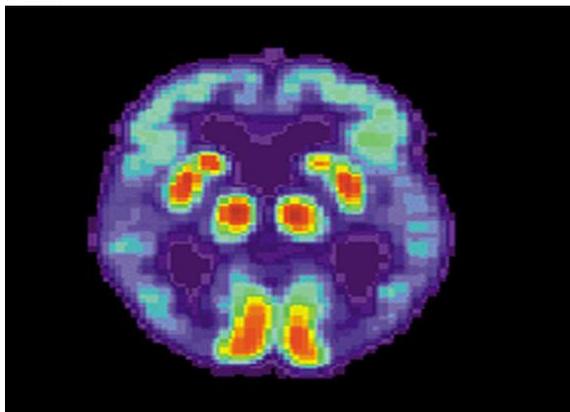


Fig 1. Otak Penderita AD

Source :
<http://www.nia.nih.gov/Alzheimers/Resources/HighRes.htm>

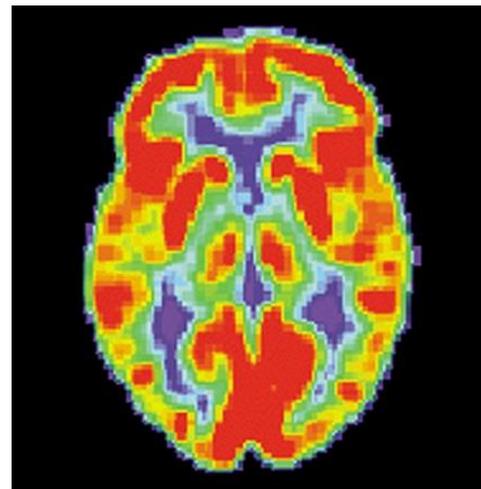


Fig 2. Otak Normal

Source :
<http://www.nia.nih.gov/Alzheimers/Resources/HighRes.htm>

Single Nucleotide Polymorphism (SNP, biasa dibaca "snip") adalah salah satu bentuk variasi materi genetik yang ditunjukkan oleh perbedaan nukleotida tunggal (adenina, timina, guanina, sitosina) di dalam susunan rangkaian basa DNA. Ini adalah salah satu bentuk paling umum dari variasi genetik manusia^[2]. Pentingnya SNPs telah direview di Genome-wide Association yang mana mempelajari hubungan SNPs dengan tanda-tanda penyakit dan metabolisme obat. Sekitar 60–90% dari variasi obat masing-masing pasien ditentukan oleh SNPs nya. Oleh karena itu, SNPs, pharmacological research, dan penemuan obat memegang peranan penting dalam mengatasi AD.

Oleh karena itu, meningkatkan informasi yang berhubungan, memerlukan bantuan dari ilmu bioinformatika untuk membuat database dan webserver yang sesuai.

Saat ini, PharmGKB (the Pharmacogenetics and Pharmacogenomics Knowledge Base) telah membangun sebuah komprehensif database yang membantu dalam mempelajari pharmacogenomic^[3]. PharmGKB menyediakan jaringan riset dalam hal penemuan SNP dan

respon obat, mencakup pengetahuan yang mendalam tentang obat, DNA, dan penyakit.

Oleh karena itu diperlukan algoritma pencarian yang tepat. Dalam makalah ini akan dijelaskan penggunaan algoritma string matching dalam proses pencarian pada database sehingga diharapkan dapat membantu riset untuk mengatasi AD.

II. THEORIES

A. DNA dan Kromosom

Asam deoksiribonukleat, lebih dikenal dengan singkatan DNA (bahasa Inggris: *deoxyribonucleic acid*), adalah sejenis biomolekul yang menyimpan dan menyandi instruksi-instruksi genetika setiap organisme dan banyak jenis virus. Instruksi-instruksi genetika ini berperan penting dalam pertumbuhan, perkembangan, dan fungsi organisme dan virus. DNA merupakan asam nukleat; bersamaan dengan protein dan karbohidrat, asam nukleat adalah makromolekul esensial bagi seluruh makhluk hidup yang diketahui.

DNA merupakan asam nukleat; bersamaan dengan protein dan karbohidrat, asam nukleat adalah makromolekul esensial bagi seluruh makhluk hidup yang diketahui. Kebanyakan molekul DNA terdiri dari dua unting biopolimer yang berpilin satu sama lainnya membentuk heliks ganda. Dua unting DNA ini dikenal sebagai polinukleotida karena keduanya terdiri dari satuan-satuan molekul yang disebut nukleotida. Tiap-tiap nukleotida terdiri atas salah satu jenis basa nitrogen (guanina (G), adenina (A), timina (T), atau sitosina (C)), gula monosakarida yang disebut deoksiribosa, dan gugus fosfat. Nukleotida-nukleotida ini kemudian tersambung dalam satu rantai ikatan kovalen antara gula satu nukleotida dengan fosfat nukleotida 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.

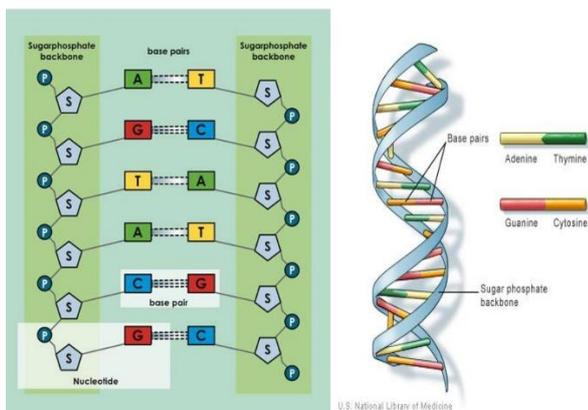


Fig 3. Struktur DNA
Source : Vivin Alfiana

Dalam sel, DNA tersusun dalam kromosom. Semasa pembelahan sel, kromosom-kromosom ini diduplikasi dalam proses yang disebut replikasi DNA. Organisme

eukariotik (hewan, tumbuhan, fungi, dan protista) menyimpan kebanyakan DNA-nya dalam inti sel dan sebagian kecil sisanya dalam organel seperti mitokondria ataupun kloroplas.^[1] Sebaliknya organisme prokariotik (bakteri dan arkaea) menyimpan DNA-nya hanya dalam sitoplasma. Dalam kromosom, protein kromatin seperti histon berperan dalam penyusunan DNA menjadi struktur kompak. Struktur kompak inilah yang kemudian berinteraksi antara DNA dengan protein lainnya, sehingga membantu kontrol bagian-bagian DNA mana sajakah yang dapat ditranskripsikan.

B. Alzheimer Disease

Penyakit alzheimer adalah hilangnya intelektual dan kemampuan bersosialisasi yang cukup parah untuk mempengaruhi aktivitas harian. Pada penyakit Alzheimer, kesehatan jaringan otak mengalami penurunan, menyebabkan menurunnya daya ingat dan kemampuan mental.

Alzheimer bukan penyakit menular, melainkan merupakan sejenis sindrom dengan apoptosis sel-sel otak pada saat yang hampir bersamaan, sehingga otak tampak mengerut dan mengecil. Alzheimer juga dikatakan sebagai penyakit yang sinonim dengan orang tua.

Alzheimer bukan merupakan bagian dari proses penuaan secara normal, akan tetapi risikonya meningkat seiring bertambahnya usia. Lima persen orang berusia di antara 65-74 tahun mengidap penyakit Alzheimer, dan hampir 50 persen orang yang berusia lebih dari 85 tahun memiliki penyakit Alzheimer.

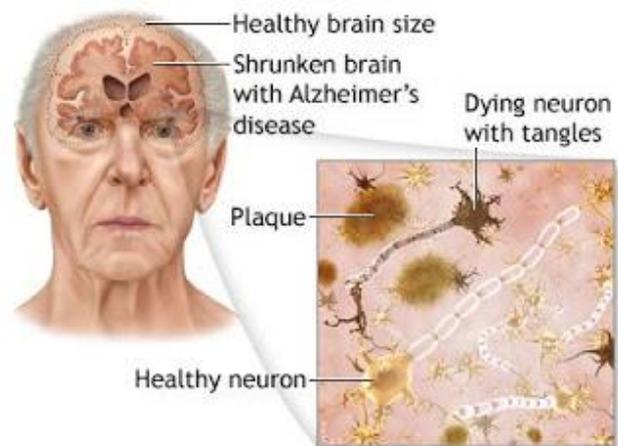


Fig 4. Gambar Otak Penderita Alzheimer
Source : Eddy Pasaribu

Penyakit Alzheimer dapat dimulai dengan hilangnya sedikit ingatan dan kebingungan, tetapi pada akhirnya akan menyebabkan pelemahan mental yang tidak dapat diubah dan menghancurkan kemampuan seseorang dalam mengingat, berpikir, belajar, dan berimajinasi. Selain itu, terdapat gejala – gejala lain, seperti :

- a. Orang dengan Alzheimer bermasalah dalam berpikir mengenai suatu hal terutama dalam bentuk angka.

- b. Sulit untuk orang dengan Alzheimer untuk menemukan kata yang tepat untuk menyampaikan pemikiran mereka atau ketika mereka terlibat pembicaraan. Pada akhirnya akan mempengaruhi kemampuan membaca dan menulis mereka.
- c. Disorientasi

C. Approximate String Matching

Dalam metode ini, dilakukan suatu penghitungan perbedaan antara dua string. Penghitungan tersebut meliputi tiga operasi string seperti di bawah ini. Untuk contoh yang akan digunakan, diasumsikan S adalah string sumber pencarian dan T adalah string yang ingin dicari.

a. Operasi Penghapusan

Misalnya S = *memori* dan T = *meri*. Penghapusan dilakukan untuk karakter m pada lokasi ke-3 dan o pada lokasi ke-4. Dua operasi penghapusan tersebut menunjukkan transformasi S ke T yang diilustrasikan sebagai berikut.

	1	2	3	4	5	6
T =	m	e	-	-	r	i
S =	m	e	m	o	r	i

b. Operasi Penyisipan

Misalnya S = *brian* dan T = *barisan*. Operasi sisip dilakukan dengan menyisipkan a dan s pada posisi 2 dan 5 yang dapat ditunjukkan sebagai berikut.

	1	2	3	4	5	6	7
T =	b	a	r	i	s	a	n
S =	b	-	r	i	-	a	n
		a			s		

c. Operasi Penggantian

Misalnya S = *perasa* dan T = *pewara*. String T ditransformasikan menjadi S dengan melakukan penggantian (substitusi) pada posisi ke-3 dan ke-5.

Huruf r dan s pada S digantikan dengan w dan r pada T. Prosesnya dapat ditunjukkan sebagai berikut.

	1	2	3	4	5	6
T =	p	e	w	a	r	a
S =	p	e	r	a	s	a
			w		r	

Ketiga operasi di atas dapat digunakan dalam contoh berikut ini. Misalnya T = *abrakadabra* dan S = *avrakadabah*. S dapat ditransformasikan menjadi T sebagai berikut.

	1	2	3	4	5	6	7	8	9	10	11	12	13
T =	a	b	r	a	k	-	a	d	a	b	r	a	-
S =	a	v	r	a	k	-	a	d	a	b	-	a	h
		s				r					w		a

Operasi yang dilakukan yaitu:

1. mengganti v pada posisi 2

2. menghapus h pada posisi 6
3. menyisipkan r pada posisi 11
4. menghapus h pada posisi 13

III. IMPLEMENTAION AND ANALYSIS

A. Implementasi

1. AD-Related Drugs Menggunakan Approximate String Matching Berdasarkan PharmGKB
Download Data

Misalkan kita menginput : “Alzheimer’s Disease “, maka akan dihasilkan tabel yang berisi kata kunci yang berhubungan dengan input.

Hasil dibawah ini merupakan kelas – kelas dimana SNP akan dicari, untuk kemudian dihasilkan obat yang sesuai.

ID Keywords

1	AD
2	Alzheimer's disease
3	AD—Alzheimer's disease
4	Acute Confusional Senile Dementia
5	Alzheimer Dementia, Presenile
6	Alzheimer Disease, Early Onset
7	Alzheimer Disease, Late Onset
8	Alzheimer Type Dementia
9	Alzheimer Type Senile Dementia
10	Alzheimer's Disease, Focal Onset
11	Alzheimer's disease, NOS
12	Dementia, Alzheimer Type
13	Dementia, Presenile
14	Dementia, Presenile Alzheimer
15	Dementia, Primary Senile Degenerative
16	Dementia, Senile
17	Dementias, Presenile
18	Dementias, Senile
19	Disease, Alzheimer
20	Disease, Alzheimer's
21	Early Onset Alzheimer Disease
22	Focal Onset Alzheimer's Disease
23	Late Onset Alzheimer Disease
24	Presenile Alzheimer Dementia
25	Presenile Dementia
26	Presenile Dementias

Fig 5 Tabel class yang berhubungan dengan "Alzheimer's Disease"

IV. KESIMPULAN

Adapun algoritma untuk mengedit jarak antara text dengan pattern pada Aproximate String Matching ditunjukkan sebagai berikut :

```

(1) // initialization
(2) for i ← 0 to m do
(3)   E(i, 0) ← i
(4) end for
(5) for j ← 0 to n do
(6)   E(0, j) ← 0
(7) end for
(8) // edit distance E(i, j)
(9) for i ← 0 to m do
(10)  for j ← 0 to n do
(11)    if(T(j) = P(i)) then
(12)      E(i, j) ← (i - 1, j - 1)
(13)    else
(14)      min ← MIN[E(i - 1, j), E(i, j - 1)]
(15)      E(i, j) ← min + 1
(16)    end if
(17)  end for
(18) end for
(19) return E

```

AD adalah bentuk demensia terbanyak pada orang tua. Pharmacogenomics dari AD masih merupakan tantangan. Pada makalah ini diharapkan membantu menemukan obat yang tepat berdasarkan gen (SNPs) menggunakan algoritma Aproximate String Matching.

Hasilnya menunjukkan dengan Aproximate String Matching, dapat dicari obat yang sesuai, bahkan informasi lain berdasarkan dengan gen pasien.

REFERENCES

- [1] M.-Y. Shiau, L. Yu, H.-S. Yuan, J.-H. Lin, and C.-K. Liu, "Functional performance of Alzheimer's disease and vascular dementia in southern Taiwan," *The Kaohsiung Journal of Medical Sciences*, vol. 22, no. 9, pp. W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
- [2] What are single nucleotide polymorphisms (SNPs)? <http://ghr.nlm.nih.gov/handbook/genomicresearch/snp> B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
- [3] K. Sangkuhl, D. S. Berlin, R. B. Altman, and T. E. Klein, "PharmGKB: understanding the effects of individual genetic variants," *Drug Metabolism Reviews*, vol. 40, no. 4, pp. 539–551, 2008. J. Wang, "Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication)," *IEEE J. Quantum Electron.*, submitted for publication.

2. Hasil PharmGKB-based data mining dalam hal PharmGKB accession ID, drug class, publications, dan informasi gen dari Alzheimer's disease.

No.	PharmGKB accession ID	Drug classes	Publications ^{*1}	Gene no. ^{*2}
1	PA164712423	Anticholinesterases	PMID:20644562 20644562 14674789	6
2	PA164712308	Ace inhibitors, plain	PMID:17362041	24
3	PA449515	Etanercept	PMID:19027875	12
4	PA451262	Rivastigmine	PMID:20644562 16323253 17082448 20644562 15289797 17522596	2
5	PA450243	Lithium	PMID:17082448	13
6	PA10384	Anti-inflammatory and antirheumatic products, nonsteroids	PMID:17082448 17082448	11
7	PA449760	Glatiramer acetate	PMID:17082448	4
8	PA133950441	Hmg coa reductase inhibitors	PMID:17082448	39
9	PA151958596	Curcumin	PMID:17082448	2
10	PA451898	Vitamin c	PMID:17082448	16
11	PA451900	Vitamin e	PMID:17082448	1
12	PA452229	Antidepressants	PMID:17082448	43
13	PA452233	Antipsychotics	PMID:17082448	46
14	PA449726	Galantamine	PMID:20644562 16323253 17082448 15853556 20644562 14674789 12177686	7
15	PA10364	Memantine	PMID:17082448	0
16	PA451283	Rosiglitazone	PMID:16770341	34
17	PA448031	Acetylcholine	PMID:15695160	8
18	PA450626	Nicotine	PMID:15695160	88
19	PA137179528	Nimesulide	PMID:16331303 11810182 PMID:20859744 20644562 16323253	3

Dapat dilihat bahwa diatas adalah daftar kelas – kelas yang berhubungan dengan “Alzheimer’s Disease”

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



Agung Baptiso Sorlawan
13513043