

Penggunaan Algoritma Pencocokkan Pola pada Aplikasi “How-Old.net”

Chairuni Aulia Nusapati 13513054
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13513054@std.stei.itb.ac.id

Abstrak—Aplikasi “How-Old.net” adalah sebuah aplikasi berbasis web baru yang dapat menebak umur dan jenis kelamin seseorang berdasarkan foto wajahnya. Aplikasi ini menggunakan algoritma deteksi wajah yang berdasarkan pencocokkan pola. Dalam makalah ini akan dianalisis mengenai cara kerja aplikasi ini dan bagaimana aplikasi dari algoritma pencocokkan pola pada aplikasi “How-old.net”.

Kata Kunci—analisis citra, aplikasi “How-Old.net”, deteksi wajah, pencocokkan pola

I. PENDAHULUAN

Baru-baru ini, muncul sebuah aplikasi berbasis web baru yang cukup menarik perhatian kalangan muda Indonesia yaitu aplikasi “How-Old.net”. Aplikasi ini dapat menerima masukan sebuah foto manusia. Kemudian aplikasi ini akan mengembalikan sebuah data yang berisi umur dan jenis kelamin dari orang-orang yang ada pada foto tersebut.

Aplikasi ini dirasa sangat menarik, karena dengan “pintar”-nya, aplikasi ini dapat menebak umur dan jenis kelamin seseorang berdasarkan fotonya. Banyak penggunanya menggunakan aplikasi ini sekedar karena penasaran apakah ia terlihat lebih muda, lebih tua, atau sama dengan umurnya yang sebenarnya,

Dibalik setiap aplikasi yang menarik, pasti ada algoritma-algoritma yang juga menarik di belakangnya. Oleh karena itu, penulis merasa tertarik untuk membahas mengenai algoritma-algoritma yang ada di belakang aplikasi yang menarik ini. Selain itu, karena penulis sedang mengikuti kuliah “Strategi Algoritma”, penulis merasa alangkah lebih baiknya jika algoritma yang dibahas adalah algoritma yang berkaitan dengan kuliah tersebut.

Setelah penulis mencari dari beberapa sumber, penul menemukan bahwa terdapat sebuah algoritma yang terdapat di balik aplikasi ini yang berkaitan dengan kuliah “Strategi Algoritma”. Algoritma tersebut adalah algoritma *pattern matching*.

II. DASAR TEORI

- A. Aplikasi “How-Old.net”
Aplikasi How-Old.net adalah sebuah aplikasi deteksi umur dan jenis kelamin yang dibuat oleh perusahaan Microsoft. Untuk menggunakan aplikasi ini, pengguna akan diminta untuk memasukkan sebuah foto manusia. Kemudian, aplikasi akan mengembalikan foto tersebut dengan “tag” yang bertuliskan umur dan jenis kelamin dari orang-orang. Tag diletakkan di wajah tiap orang pada foto. Masyarakat dapat mengakses aplikasi ini pada situs <http://how-old.net/>.

Berikut gambar-gambar yang bersangkutan dengan penggunaan aplikasi ini.



Gambar 1: Halaman utama aplikasi “How-Old.net”

Sumber: <http://how-old.net>

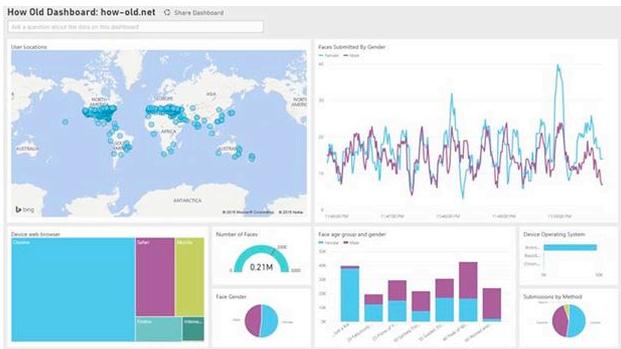


Gambar 2: Hasil analisis foto

Sumber: <http://how-old.net/#results>

Awalnya, program ini dibuat sebagai aplikasi percobaan untuk mengenalkan kepada publik betapa mudahnya membuat aplikasi “pintar” menggunakan *Azure services* buatan Microsoft pada sebuah konferensi *Microsoft's Build2015 developer*. Peluncuran aplikasi ini juga awalnya hanya sebagai Beta test untuk menguji keakuratan algoritmanya. Para pembuatnya hanya berharap setidaknya beberapa puluh orang akan mencoba aplikasi ini dan membantunya menjadi lebih baik. Namun, ternyata setelah aplikasi ini diluncurkan pada awal 2015, aplikasi ini mendapat respon publik yang luar biasa. Banyak orang menggunakan aplikasi ini untuk mendeteksi bagaimana wajahnya terlihat. Apakah lebih tua, lebih muda, atau sesuai dengan umurnya yang sebenarnya. Dalam tiga jam pertama, aplikasi ini telah digunakan sebanyak 210.000 kali. Angka yang cukup fantastis jika dibandingkan dengan ekspektasi awal para pembuatnya. Hingga sekarang, aplikasi ini telah menjadi sebuah aplikasi yang populer. Juga, walaupun hasilnya terkadang tidak tepat, aplikasi ini adalah aplikasi yang menarik dan mengundang gelak tawa.

Berikut gambar yang data yang menggambarkan respon publik atas aplikasi ini setelah tiga jam diluncurkan.



Gambar 3: Data respon publik atas aplikasi “How-Old.net” setelah tiga jam diluncurkan

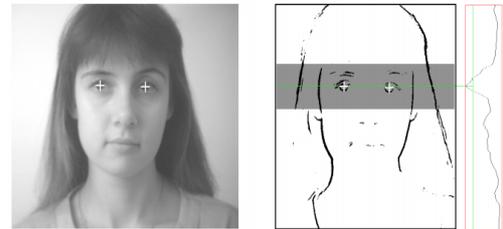
Aplikasi ini dibuat menggunakan API baru yaitu *Face detection API* dan kakas *Azure service* milik Microsoft. *Face detection API* adalah sebuah API yang dapat digunakan untuk membuat aplikasi-aplikasi yang berhubungan dengan deteksi wajah. Dasar dari API ini adalah face detection algorithm. Algoritma face detection akan dibahas pada subbab berikutnya. Kakas *Azure Service* adalah kakas yang mempermudah pembuatan aplikasi “pintar”.

B. Face Detection Algorithm

Face Detection Algorithm adalah algoritma yang

dapat digunakan untuk mengenali pola-pola pada wajah. Salah satu dasar dari algoritma ini adalah analisis citra, yang merupakan bagian dari aplikasi algoritma stringmatching.

Berikut gambar yang mendeskripsikan deteksi wajah.



Gambar 4: Analisis citra wajah berdasarkan letak mata

Sumber:

http://www.handysolution.com/science/Face_rec1_2.pdf



Gambar 5: Gambar foto sebenarnya (atas) dan gambar olahan citranya (bawah), masing-masing terurut atas dan bawah.

Terdapat beberapa jenis algoritma face detection. Namun, idenya hampir semua sama. Algoritma ini pertama-tama akan mencari beberapa hal tertentu yang dapat mencirikan wajah tersebut. Hal-hal ini kemudian akan membentuk pola wajah. Contoh hal yang dapat mencirikan wajah adalah letak garis kerutan, perbandingan jarak kedua mata dan jarak kening ke hidung, dan dalam kantung mata. Kemudian, pola ini kemudian akan digunakan untuk analisis selanjutnya.

Analisis selanjutnya dapat bermacam-macam tergantung tujuan dari face detection algorithm yang dipakai. Jika tujuannya hanya untuk menentukan apakah itu wajah manusia atau bukan,

maka cukup dengan memastikan bahwa pola wajah tersebut sudah cocok dengan pola wajah manusia, maka analisis akan memberikan jawaban. Namun, jika digunakan untuk menebak sebuah karakter yang berasal dari kombinasi beberapa bagian dari pola seperti penentuan umur, maka pola harus dicocokkan satu persatu dengan pola yang ada pada penentuan umur.

Pencocokkan pola pada algoritma ini pada umumnya menggunakan algoritma pencocokkan pola/string yang sudah ada dan dipelajari.

C. Algoritma String/Pattern Matching

Pencocokan Untai/Pola adalah pencarian sebuah unta/pola tertentu dalam sebuah teks. Jika sebuah unta/pola tersebut ditemukan pada sebuah teks, maka pencocokkan unta/pola tersebut dinyatakan berhasil dan lokasi penemuan unta/pola pertama pada teks akan menjadi jawaban dari persoalan ini.

Berikut definisi pencocokkan unta/pola.

Diberikan:

T: teks (text), yaitu (long) string yang panjangnya n karakter

P: pattern, yaitu string dengan panjang m karakter (asumsi $m < n$) yang akan dicari di dalam teks.

Carilah (find atau locate) lokasi pertama di dalam teks yang bersesuaian dengan pattern.

Contoh penggunaan:

T: "the rain in spain stays mainly on the plain"

P: "main"

Dalam dunia nyata, algoritma ini banyak digunakan untuk menyelesaikan persoalan yang membutuhkan pencocokkan pola. Persoalan-persoalan tersebut contohnya adalah pencarian pada editor teks, pencarian pada web search engine, analisis sidik jari, pencocokkan rantai asam amino pada DNA, dan analisis wajah. Pada makalah ini aplikasi string matching yang akan dibahas adalah pada analisis wajah (analisis citra pada wajah).

Beberapa contoh algoritma pencocokan unta/pola yang dikenal umum dan diajarkan dalam mata kuliah "Strategi Algoritma" adalah algoritma Brute Force, Boyer-Moore, dan Knuth-Pratt-Morris. Namun, karena performansinya yang lebih baik, algoritma Boyer-Moore dan Knuth-Pratt-Morris lebih umum digunakan. Berikut penjelasan mengenai masing-masing Algoritma.

1. Algoritma Brute Force

Algoritma ini melakukan pencocokkan dari kiri ke

kanan. Algoritma ini akan melakukan penelusuran mulai dari indeks pertama T hingga indeks terakhir teks untuk menentukan apakah P terletak mulai pada indeks tersebut. Apabila didapat bahwa benar P terletak mulai pada indeks tersebut, algoritma akan melanjutkan pencocokkan untuk indeks selanjutnya (kedua) dari P dan seterusnya. Apabila sepanjang P tersebut cocok, maka pola telah ditemukan.

2. Algoritma Knuth-Pratt-Morris

Algoritma ini merupakan perbaikan dari algoritma Brute Force. Pada brute force, penelusuran berikutnya selalu dilakukan pada indeks tepat setelahnya. Namun, pada algoritma ini pencocokkan berikutnya tidak selalu dilakukan pada indeks tepat setelahnya atau dalam kata lain, pergeseran hanya dilakukan sebanyak satu karakter. Namun, pergeseran dapat dilakukan sebanyak lebih dari satu karakter. Berikut penjelasan mengenai pergeseran tersebut.

Jika didapat bahwa terjadi ketidakcocokkan karakter antara T dan P pada $P[j]$, pergeseran yang dilakukan adalah sebesar panjang prefix terbesar yang juga merupakan suffix dari $P[1..j-1]$. Jika panjang terbesarnya adalah 0 atau 1 , maka pergeseran dilakukan sebanyak satu karakter saja. Pada aplikasinya, perhitungan semua panjang prefix yang juga merupakan suffix dari setiap substring $P[1..i]$ dilakukan di awal program dan akan dimasukkan ke dalam sebuah tabel, kemudian program cukup mengakses tabel ini. Hasil perhitungan pada tabel tersebut disebut dengan border function $b(i)$. $b(i)$ menyatakan panjang prefix yang juga merupakan suffix dari setiap substring $P[1..i]$, dan hasil ini akan disimpan di tabel pada indeks ke- i .

3. Algoritma Boyer-Moore

Algoritma ini agak berbeda dari algoritma lainnya. Pada algoritma-algoritma sebelumnya, indeks pertama P adalah indeks yang pertama dicocokkan. Sementara, pada algoritma ini indeks terakhir P adalah indeks yang pertama dicocokkan dan pencocokkan pada P dilakukan dari kanan ke kiri. Namun, untuk pergeseran P tetap dilakukan dari kiri ke kanan. Pada kasus tertentu, dapat dilakukan character jumping yaitu pergeseran yang cukup jauh (lebih dari satu). Berikut penjelasan mengenai character-jumping.

Jika pada penelusuran, $P[i] \neq T[i]$ terjadi, maka terdapat tiga kemungkinan character-jump. Misalkan karakter $T[i]$ adalah x . Pertama, jika karakter x terdapat pada $P[i]$, maka P akan digeser ke kanan agar karakter x pada T sejajar dengan karakter x yang muncul pertama kali pada P .

Kedua, jika karakter x terdapat pada P[i] namun menyejajarkan karakter x pada T dengan x yang muncul pertama kali pada P membuat P harus bergerak ke kiri, karakter pertama P akan disejajarkan dengan T[i+1]. Kasus terakhir, jika kedua kasus tadi tidak berlaku, maka akan dilakukan pergeseran sebesar satu karakter.

Perhitungan apakah setiap karakter yang dapat muncul pada T telah muncul pada P dilakukan di awal algoritma. Kemudian, algoritma cukup mengakses sebuah tabel yang berisi jumlah kemunculan pada P dari tiap karakter yang dapat muncul pada T.

III. ANALISIS

A. Cara Kerja Aplikasi “How-Old.net”

Aplikasi “How-Old.net” adalah aplikasi yang berdasarkan pada algoritma face detection, sebuah analisis citra. Analisis citra adalah analisis yang didasarkan pada analisis pola yang ada pada citra yang diolah.

Pada aplikasi “How-Old.net”, sebuah foto yang masuk akan dianalisis citranya, menghasilkan sebuah pola yang disimpan pada program. Kemudian, pola ini akan dianalisis kecocokannya dengan pola yang sudah ada. Namun, analisis kecocokan pola pada aplikasi ini tidak serta merta langsung dicocokkan dengan pola yang ada. Pencocokan dilakukan satu persatu untuk tiap parameter analisis.

Setelah pola dicocokkan untuk tiap parameter analisis, hasil analisis ini akan menjadi data yang dapat digunakan untuk menganalisis umur dan jenis kelamin orang yang ada pada foto.

B. Penggunaan Pattern Matching Algorithm pada Aplikasi “How-Old.net”

Pada tahap analisis citra dan analisis pola yang dijelaskan pada subbab sebelumnya, terjadi pencocokkan pola. Pencocokkan pola ini menggunakan algoritma pencocokkan pola yang telah ada dan dikenal.

Ilustrasinya adalah sebagai berikut. Misalkan program sedang akan menganalisis umur seorang yang kita beri nama “Sanchez”. Aplikasi akan menganalisa beberapa parameter yang ada di wajah Sanchez. Dalam pembahasan ini, dimisalkan bahwa parameter yang akan digunakan untuk analisis adalah kerutan di dahi, kerutan di sekitar bibir, dan kantung mata.

Pertama, program akan menganalisis kerutan dahi Sanchez dengan mencocokkan pola pada kerutan dahinya. Maka program akan membuat

pola dari kerutan dahi tersebut. Misalkan pola yang terbentuk adalah “Dcd3 5Ac d 3 5 x 7 I 2 3 Z”. Dengan tiap pemisahan 4 digit mengartikan kerutan di dahi kanan, dahi atas, dahi kiri, dan dahi bawah secara sirkular. Misalkan kode untuk kerutan berat adalah “cd35”. Sehingga jika dalam pola kerutan di dahi tadi terdapat n buah kecocokan pola maka terdapat n buah kerutan berat. Misalkan juga diketahui dari teori, bahwa jika terjadi dua kerutan berat pada dahi maka orang tersebut berusia >50 tahun, jika hanya satu berusia 30-50 tahun, jika tidak ada berusia <30 tahun.

Pada tahap ini, akan dilakukan pencocokkan pola pada pola kerutan di dahi Sanchez tadi.

Dalam makalah ini, akan digunakan pendekatan tiga buah algoritma untuk mencocokkannya, untuk meningkatkan akurasi. Walaupun sebenarnya dalam aplikasi ini hanya digunakan sebuah algoritma yang dipilih pengembangnya.

1. Algoritma Brute Force

D	c	d	3	5	A	c	d	3	5	x	7	I	2	3	Z
c	d	3	5												
	c	d	3	5											
		c	d	3	5										
			c	d	3	5									
				c	d	3	5								
					c	d	3	5							
						c	d	3	5						
							c	d	3	5					
								c	d	3	5				
									c	d	3	5			
										c	d	3	5		

Jumlah kecocokkan: 2

2. Algoritma BM

D	c	d	3	5	A	c	d	3	5	x	7	I	2	3	Z
c	d	3	5												
	c	d	3	5											
		c	d	3	5										
			c	d	3	5									
				c	d	3	5								
					c	d	3	5							
						c	d	3	5						
							c	d	3	5					
								c	d	3	5				
									c	d	3	5			

Jumlah kecocokkan: 2

3. Algoritma KMP

D	c	d	3	5	A	c	d	3	5	x	7	I	2	3	Z
c	d	3	5												
	c	d	3	5											
		c	d	3	5										
			c	d	3	5									
				c	d	3	5								
					c	d	3	5							
						c	d	3	5						
							c	d	3	5					
								c	d	3	5				
									c	d	3	5			

				c	d	3	5												
					c	d	3	5											
						c	d	3	5										
							c	d	3	5									
								c	d	3	5								
									c	d	3	5							
										c	d	3	5						
											c	d	3	5					
												c	d	3	5				

Jumlah kecocokkan: 2

Berdasarkan komputasi tiga algoritma di atas, didapat bahwa pada dahi Sanchez terdapat dua buah kerutan besar. Ini menyimpulkan bahwa Sanchez berumur >50 tahun.

Kemudian, program akan menganalisis kerutan di sekitar bibir Sanchez. Misalkan cara analisis kerutan di sekitar bibir serupa dengan analisis kerutan di dahi, namun perbedaannya terletak pada pola yang dianalisis.

Kemudian, program akan menganalisis kerutan di sekitar bibir Sanchez dengan mencocokkan pola pada kerutan di sekitar bibirnya. Program akan membuat pola dari kerutan dahi tersebut. Misalkan pola yang terbentuk adalah "B2Z69c7g2odfccxz". Misalkan kode untuk kerutan berat adalah "9c7g". Sehingga jika dalam pola kerutan di sekitar bibir tadi terdapat n buah kecocokkan pola maka terdapat n buah kerutan berat. Misalkan juga diketahui dari teori, bahwa jika terjadi dua kerutan berat pada sekitar bibir seseorang maka orang tersebut berusia >60 tahun, jika hanya satu berusia 30-60 tahun, jika tidak ada berusia <30 tahun. Berikut pencocokkan polanya.

1. Algoritma Brute Force

B	2	Z	6	9	c	7	g	2	o	d	f	c	c	x	z
9	c	7	g												
	9	c	7	g											
		9	c	7	g										
			9	c	7	g									
				9	c	7	g								
					9	c	7	g							
						9	c	7	g						
							9	c	7	g					
								9	c	7	g				
									9	c	7	g			
										9	c	7	g		
											9	c	7	g	

Jumlah kecocokkan: 1

2. Algoritma BM

B	2	Z	6	9	c	7	g	2	o	d	f	c	c	x	z
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

9	c	7	g																
				9	c	7	g												
								9	c	7	g								
												9	c	7	g				

Jumlah kecocokkan: 1

3. Algoritma KMP

B	2	Z	6	9	c	7	g	2	o	d	f	c	c	x	z
9	c	7	g												
	9	c	7	g											
		9	c	7	g										
			9	c	7	g									
				9	c	7	g								
					9	c	7	g							
						9	c	7	g						
							9	c	7	g					
								9	c	7	g				
									9	c	7	g			
										9	c	7	g		
											9	c	7	g	
												9	c	7	g

Jumlah kecocokkan: 1

Berdasarkan komputasi tiga algoritma di atas, didapat bahwa pada sekitar bibir Sanchez terdapat satu buah kerutan besar. Ini menyimpulkan bahwa Sanchez berumur 30-60 tahun.

Terakhir, program akan menganalisis kantung mata Sanchez. Misalkan dari kantung mata Sanchez didapat pola "08xx16h72n34". Misalkan kode untuk garis hitam dalam adalah "2n34". Sehingga jika dalam pola kantung mata tadi terdapat n buah kecocokkan pola maka terdapat n buah garis hitam dalam. Misalkan juga diketahui dari teori, bahwa jika terjadi dua orang tersebut berusia >60 tahun, jika hanya satu berusia 40-60 tahun, jika tidak ada berusia <40 tahun. Berikut pencocokkan pola kantung mata Sanchez.

1. Algoritma Brute Force

0	8	x	x	1	6	h	7	2	n	3	4				
2	n	3	4												
	2	n	3	4											
		2	n	3	4										
			2	n	3	4									
				2	n	3	4								
					2	n	3	4							
						2	n	3	4						
							2	n	3	4					
								2	n	3	4				
									2	n	3	4			
										2	n	3	4		
											2	n	3	4	
												2	n	3	4

Jumlah kecocokkan: 1

2. Algoritma BM

