

Algoritma pencocokan String dalam Bidang Forensik

Ade Surya Handika 13518007
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13518007@std.stei.itb.ac.id

Abstract—Forensik adalah ilmu yang menerapkan berbagai macam lintas ilmu untuk melakukan adanya tindakan hukum. Salah satu metode yang dilakukan dalam forensik yaitu dengan pencocokan DNA. DNA terdiri dari empat basa nitrogen Adenin (A), Guanin (G), Sitosinin (C), dan Timin(T). Suatu rangkaian DNA dapat dianggap sebagai sebuah string yang memiliki karakter (A,G,C,T). Dengan demikian, pencocokan DNA dapat dilakukan dengan algoritma pencocokan String. Algoritma pencocokan string yang digunakan dapat bermacam-macam, misalnya brute force, knuth morris, boyer moore. Dengan memanfaatkan algoritma pencocokan string maka pola suatu DNA yang ditemukan ketika identifikasi forensik pada suatu kasus dapat diketahui siapa yang memiliki DNA dengan pola yang sama dengan yang ditemukan dan hal ini dapat dijadikan bukti pada penegakan hukum.

Keywords—Forensik, Pencocokan DNA, Algoritma Pencocokan String.

I. PENDAHULUAN

Forensik adalah sebuah penerapan dari lintas ilmu pengetahuan untuk menjawab berbagai pertanyaan yang penting untuk sebuah sistem hukum. Ilmu forensik sering dikaitkan dengan adanya analisis atas tindakan hukum. Namun, dengan mengesampingkan keterkaitannya dengan sistem hukum, forensik secara umum lebih meliputi sesuatu atau metode-metode yang bersifat ilmiah (bersifat ilmu) dan juga aturan-aturan yang dibentuk dari fakta-fakta berbagai kejadian, untuk melakukan pengenalan terhadap bukti-bukti fisik. Contoh dari bukti-bukti fisik ini seperti mayat, bangkai, dll. Salah satu cara atau metode yang digunakan dalam forensik dalam menganalisis yaitu dengan pencocokan DNA dan Pemeriksaan Sidik Jari. Kedua hal ini dapat dilakukan untuk mengidentifikasi bukti fisik yang ada.

Pencocokan DNA dapat dilakukan dengan menggunakan algoritma pencocokan String. Algoritma pencocokan string adalah algoritma yang digunakan untuk mencari sebuah pola dalam suatu string panjang. Algoritma pencocokan string memiliki banyak variasi, seperti brute force, Knuth Morris dan Boyer Moore. Dengan brute force algoritma akan mengecek suatu karakter dengan menggeser satu persatu pada string dan membandingkannya dengan karakter pada pola yang ada.

Algoritma Knuth-Morris hampir sama seperti brute force, hanya pada saat pergeseran karakter dilakukan dengan lebih cerdas. Sedangkan Algoritma Boyer Moore agak berbeda dari contoh sebelumnya, yaitu dilakukan teknik mirror-looking glass. Dengan memanfaatkan algoritma pencocokan string, bukti forensik seperti potongan dna dapat dicocokkan dengan data yang dimiliki untuk mengidentifikasi pelaku yang ada pada kasus forensik.

II. LANDASAN TEORI

2.1 Forensik

2.1.1 Pengertian Ilmu Forensik

Forensik (berasal dari bahasa Yunani 'Forensis' yang berarti debat atau perdebatan) adalah bidang ilmu pengetahuan yang digunakan untuk membantu proses penegakan keadilan melalui proses penerapan ilmu (sains)¹. Dalam kelompok ilmu-ilmu forensik ini dikenal antara lain ilmu fisika forensik, ilmu kimia forensik, ilmu psikologi forensik, ilmu kedokteran forensik, ilmu toksikologi forensik, komputer forensik, ilmu balistik forensik, ilmu metalurgi forensik dan sebagainya.

ilmu forensik menerapkan ilmu pengetahuan dengan tujuan penetapan hukum dan pelaksanaan hukum dalam sistem peradilan hukum pidana maupun hukum perdata.

2.1.2 Metode Forensik

2.1.2.1 Pemeriksaan Sidik Jari

Metode ini membandingkan data sidik jari jenazah dengan data sidik jari antemortem. Metode ini diakui memiliki ketepatan yang tinggi. (Budiyanto, 1997)

2.1.2.2 Pemeriksaan DNA

Gambaran DNA tiap individu sangat spesifik dan dapat dijadikan patokan dalam identifikasi. Kemungkinan dua individu yang tidak memiliki hubungan darah untuk memiliki sekuens DNA yang sama sangat kecil yaitu 1 : 1.000.000.000. (James, 2011)

2.1.2.3 Pemeriksaan Gigi

Pemeriksaan ini membandingkan data gigi dan rahang jenazah dengan data antemortem. Seperti sidik jari, data gigi setiap individu juga berbeda satu sama lainnya. Pencatatan data

gigi dan rahang (odontogram) dilakukan secara manual, sinar-X, dan pencetakan gigi dan rahang. Data ini berisi tentang jumlah gigi, susunan, bentuk, tambalan, gigi palsu dan sebagainya. (Budiyanto, 1997)

2.1.2.4 Metode Visual

Metode ini dilakukan dengan memperlihatkan jenazah kepada orang-orang yang merasa kehilangan kerabatnya. Jenazah sebaiknya dalam keadaan yang belum membusuk sehingga wajah dan bentuk tubuh masih dapat dikenali oleh lebih dari satu orang. Metode ini juga harus memperhatikan faktor emosi kerabat dalam membenarkan atau menyangkal identitas jenazah tersebut. (Budiyanto, 1997)

2.1.2.5 Pemeriksaan Dokumen

Metode ini dapat dilakukan apabila ditemukan dokumen yang berisikan identitas seperti kartu identitas pribadi, surat izin mengemudi dan sebagainya di dalam saku pakaian yang dikenakan jenazah. Dokumen yang berada di dekat jenazah belum tentu merupakan milik jenazah yang bersangkutan, terutama pada kasus seperti kecelakaan masal. (Budiyanto, 1997)

2.1.2.6 Pemeriksaan Pakaian dan Perhiasan

Pada pakaian dan perhiasan yang dikenakan jenazah kemungkinan dapat diperoleh data berupa merk pakaian, ukuran, inisial nama, lencana dan sebagainya yang dapat membantu, walaupun telah terjadi proses pembusukan pada jenazah tersebut. (Budiyanto, 1997)

2.1.2.7 Identifikasi Medik

Metode ini menggunakan data tinggi badan, berat badan, warna rambut, warna mata, kelainan khusus, tato dan sebagainya. Ketepatan metode ini cukup tinggi karena dilakukan seorang ahli dengan menggunakan beberapa cara. (Budiyanto, 1997)

2.1.2.8 Pemeriksaan Serologik

Pemeriksaan ini bertujuan untuk menentukan golongan darah jenazah. Pada jenazah yang sudah membusuk dapat diperiksa dari rambut, kuku dan tulang. (Budiyanto, 1997)

2.1.2.9 Metode Eksklusi

Metode ini digunakan pada kecelakaan massal yang melibatkan sejumlah korban yang data identitasnya dapat diketahui, seperti penumpang pesawat dan kapal laut. Metode ini dilakukan terhadap sisa korban yang tidak dapat diidentifikasi dengan metode lain. (Budiyanto, 1997)

2.1.2.10 Identifikasi Potongan Tubuh Manusia

Pemeriksaan ini bertujuan untuk menentukan apakah potongan tubuh yang ditemukan merupakan potongan tubuh manusia dan apakah berasal dari satu tubuh atau tidak. Penentuan potongan tubuh manusia dilakukan dengan pemeriksaan jaringan makroskopik, mikroskopik dan

serologik. Selain itu dari pemeriksaan juga ditentukan jenis kelamin, ras, perkiraan umur, tinggi badan dan keterangan lain. (Budiyanto, 1997)

2.1.2.11 Identifikasi Kerangka

Identifikasi ini bertujuan untuk menentukan apakah itu kerangka manusia, jenis kelamin, ras, perkiraan umur, tinggi badan, deformitas, tanda kekerasan dan perkiraan saat kematian. (Budiyanto, 1997)

2.2 DNA

2.2.1. Pengertian DNA

Asam deoksiribonukleat, lebih dikenal dengan DNA (deoxyribo nucleic acid), adalah sejenis asam nukleat yang tergolong biomolekul utama penyusun organisme. DNA mengandung informasi yang ada pada suatu organisme. DNA ini dapat diwariskan oleh induk kepada keturunannya. DNA bertugas menginstruksikan pembentukan protein yang mana protein tersebut berperan dalam pertumbuhan, perkembangan, reproduksi, fungsi dan perilaku dari organisme.

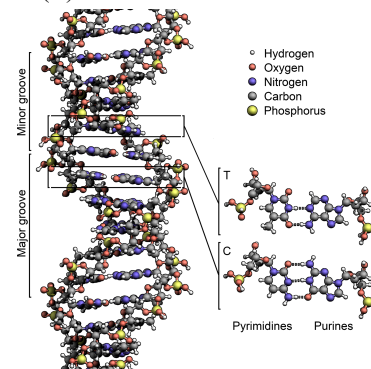
2.2.2. Struktur DNA

Struktur kimia dari suatu DNA berupa makromolekul kompleks yang terdiri dari 3 macam molekul yaitu :

- gula pentosa (deoksiribosa)
- asam fosfat
- basa nitrogen.

Pada DNA, basa nitrogen yang ditemukan adalah

- Adenina (A)
- Guanina (G)
- Sitosina (C)
- Timina (T)



gambar 2.1 Struktur DNA, sumber: wikipedia

Basa nitrogen akan saling berpasang-pasangan di antara dua nukleotida hingga terlihat seperti anak tangga. Adenine (A) akan berpasangan dengan Thymine (T) sedangkan Sitosin (C) akan berpasangan dengan Guanin (G). Sementara itu, tugas memasangkan basa nitrogen dilakukan oleh deoksiribosa dan fosfat.² Bentuk dari sebuah DNA adalah heliks ganda

2.2.3 DNA pada bidang Forensik

Dalam forensik DNA umumnya didapat dari sampel darah, sperma, kulit, liur atau rambut yang tersisa di tempat kejadian kejahatan untuk mengidentifikasi kemungkinan tersangka. Proses ekstraksi ini fingerprinting genetika atau pemprofilan DNA (DNA profiling). Proses mulai dari pengambilan sampai hingga pemprofilan DNA umumnya sangat lama sehingga terkadang pihak berwajib membutuhkan terdakwa dari kejahatan tertentu untuk menyediakan sebuah contoh DNA untuk dimasukkan ke dalam basis data komputer.

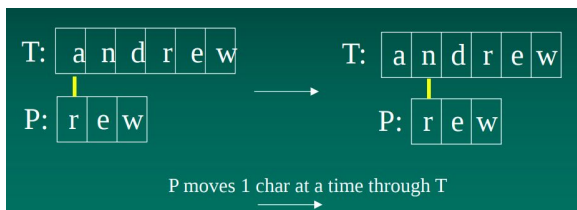
2.3 Algoritma Pencocokan String

2.3.1 Definisi Algoritma Pencocokan String

Algoritma pencocokan string adalah algoritma yang digunakan untuk mencari sebuah pola pada String. Dalam pencocokan string, terdapat sebuah teks (T) dan pola (P). Untuk suatu karakter pada teks didefinisikan dengan $T[i]$ yang mana i adalah posisi karakter pada teks.

2.3.2 Algoritma Brute Force

Pada Algoritma Brute Force pengecekan dilakukan dengan mengecek untuk setiap karakter pada T apakah ada di dalam P dimulai dari posisi awal P. Ketika terjadi mismatch atau ketidaksesuaian maka akan dilakukan pergeseran untuk karakter setelahnya untuk di cek lagi. Kompleksitas algoritma ini adalah $O(m \times n)$ dimana m adalah panjang teks dan n adalah panjang pola.



gambar 2.2 Visualisasi pengecekan menggunakan algoritma Brute force, sumber:

[https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-\(2018\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-(2018).pdf)

2.3.3 Algoritma Knuth-Morris Pratt

Algoritma Knuth-Morris Pratt adalah algoritma pencocokan string yang dimulai pengecekan dari karakter awal hingga akhir layaknya brute force. Hanya saja, pergeseran karakter untuk pengecekan selanjutnya ketika pola tidak cocok berbeda dari brute force dan lebih efisien.

Pada kasus ketika $P[j]$ tidak sama dengan $T[i]$ pergeseran tidak lagi dilakukan satu kali, melainkan dengan melihat nilai panjang prefix pada Pola yang sama dengan postfix pada Pola.

Nilai ini disimpan pada sebuah tabel dan dihitung dengan fungsi yang dinamakan border function.

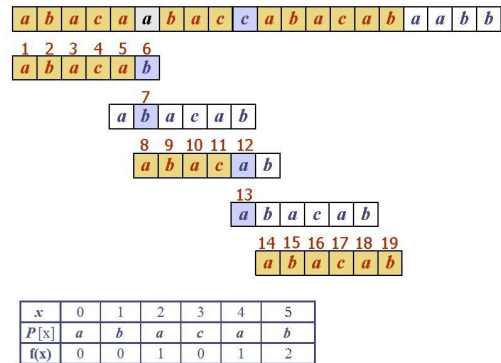
Kompleksitas algoritma ini adalah $O(m + n)$

2.3.3.1 Border Function (b)

Border Function adalah fungsi yang akan memetakan nilai terpanjang prefix dari $P[0 \dots j-1]$ yang sama dengan postfix $P[1 \dots j-1]$ berdasarkan nilai yang di input pada fungsi. Border function akan bertindak layaknya tabel.

Contohnya untuk Teks: aabaa dan Pola: ab

Ketika memanggil border function $b(2)$ maka akan di cek nilai dari terpanjang prefix $P[0..1]$ yang sama dengan postfix $P[1..1]$ yang mana hasilnya adalah 1 yaitu karakter b.



gambar 2.3 Visualisasi pengecekan menggunakan algoritma Knuth Morris Pratt, sumber: i-satyam.blogspot.com

2.3.4 Algoritma Boyer Moore

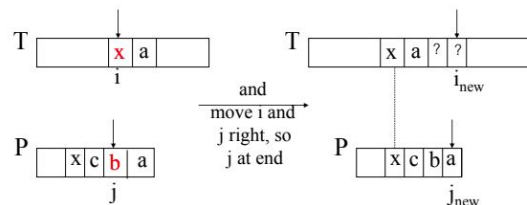
Algoritma boyer moore merupakan algoritma yang paling mangkus untuk pencocokan string. Algoritma ini banyak digunakan dalam teks editor seperti microsoft word untuk 'find and replace'

Algoritma boyer moore dilakukan dengan dua teknik

- Pertama yaitu *looking-glass technique*, merupakan teknik mencari P pada T dengan memulai dari bagian akhir P.
- Kedua yaitu *character jump technique*, yaitu teknik yang dilakukan ketika terjadi ketidakcocokkan $T[i]$ dengan x dan karakter $T[i]$ tidak sama dengan $P[j]$. x adalah suatu karakter.

Pada *character jump*. dapat terjadi 3 macam kasus

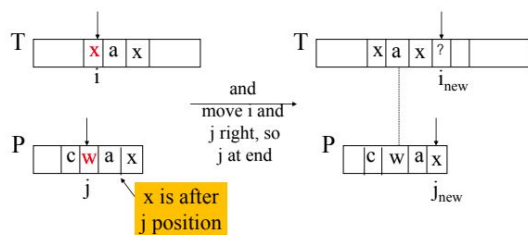
Kasus pertama ketika x terdapat di P kemudian dilakukan pergeseran kekanan untuk mensejajarkan x terakhir kali ditemukan pada P.



gambar 2.3 Visualisasi kasus pertama jump technique sumber:

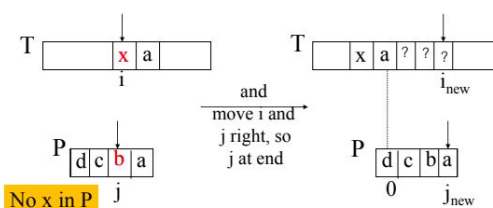
[https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-\(2018\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-(2018).pdf)

Kasus kedua ketika x terdapat di P namun tidak memungkinkan untuk melakukan pergeseran kekanan sehingga P digeser kekanan satu langkah pada T.



gambar 2.4 Visualisasi kasus kedua jump technique, sumber: [https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-\(2018\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-(2018).pdf)

Kasus ketika yaitu ketika kasus pertama dan kedua tidak bisa lakukan maka sejajarkan P[0] dengan T[i+1]



gambar 2.5 Visualisasi kasus ketika jump technique sumber:

[https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-\(2018\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-(2018).pdf)

2.3.4.1 Last Occurrence Function

Pada algoritma boyer moore terdapat preprocessing dengan mencari nilai last occurrence suatu karakter x yang ada pada pola dengan fungsi last occurrence L(). Fungsi ini akan memetakan semua karakter A menjadi nilai yang menunjukkan indeks prefix P[i] yang sama dengan x. Nilai fungsi akan -1 jika x tidak ada dalam P.

Contoh untuk pola "abacab" mengandung himpunan karakter (A, B, C) dan A adalah himpunan karakter (A, B, C, D) maka nilai L(A) adalah 4 karena karakter A ditemukan pada indeks ke 4 pada pola. Indeks awal bernilai 0.

Fungsi last occurrence biasanya akan menyimpan senarai yang berisi nilai last occurrence dengan suatu karakter x sebagai key dari senarai tersebut.

III. PEMBAHASAN

3.1 Implementasi Algoritma Brute Force untuk Pencocokan DNA

Seperti yang sudah disebutkan sebelumnya bahwa Pencocokan DNA merupakan salah satu metode forensik dan DNA dapat dianggap sebagai suatu rangkaian string sehingga pencocokan DNA dapat dilakukan dengan algoritma pencocokan string.

Berikut contoh aplikasi algoritma brute force:

Suatu rangkaian DNA (T)
GCATAGCAGAGAGTATAC

akan dicocokkan dengan potongan atau pola DNA (P)
CAG

Langkah Pertama:

Melakukan pencocokan karakter karakter awal pada DNA (T[0]) dengan karakter awal pada pola (P[0])

GCATAGCAGAGAGTATAC
CAG

Karena karakter tidak sama maka dilakukan pergeseran pola satu langkah kekanan DNA.

GCATAGCAGAGAGTATAC
CAG

Langkah Kedua:

Karakter pertama pada pola sama dengan karakter teks, dengan demikian lakukan pengecekan karakter selanjutnya pada pola dengan karakter T[i+1]. Ulangi sampai karakter terakhir

GCATAGCAGAGAGTATAC
CAG

Ternyata karakter terakhir pada pola berbeda dengan karakter yang dicocokkan sehingga dilakukan pergeseran kembali

Langkah Ketiga:

Pencocokan karakter terus dilakukan secara berulang seperti langkah sebelumnya sehingga didapatkan posisi berikut

GCATAGCAGAGAGTATAC
CAG

Saat dilakukan pencocokan, semua karakter pada pola sama dengan karakter pada DNA sehingga Algoritma berhenti dan pola CAG ada dalam pola GCATAGCAGAGAGTATAC

Kasus Ketika Tidak Ditemukan:

Ketika pencocokan karakter sudah melebihi panjang Teks yang dicocokkan, yaitu ketika panjang pola kurang dari sisa karakter yang dapat dicocokkan maka pola dianggap tidak ditemukan.

GCATAGCAGAGAGTATAC
CGA

IV. KESIMPULAN

Aplikasi algoritma pencocokan string memiliki banyak sekali manfaatnya. Salah satunya dalam bidang forensik. Analisis forensik dengan pencocokan DNA terkadang dilakukan. Dengan menganggap sebuah rangkaian DNA manusia sebagai sebuah teks dan DNA yang ditemukan ketika dilakukan forensik sebagai pola DNA maka dapat dilakukan pencocokan dengan algoritma pencocokan string.

IV. UCAPAN TERIMA KASIH

Penulis bersyukur atas nikmat Allah SWT yang telah memberikan kesehatan dan keselamatan sehingga penulis mampu menyelesaikan makalah ini dengan tepat waktu. Penulis juga ingin berterima kasih kepada seluruh dosen pengampu mata kuliah Strategi Algoritma yang telah memberikan banyak ilmu kepada penulis. Penulis juga menyampaikan terima kasih kepada teman-teman penulis dan juga semua orang yang telah membantu dalam penyelesaian makalah ini.

REFERENCES

- [1] <http://ondigitalforensics.weebly.com/digital-forensics>
- [2] <https://doktersehat.com/dna>
- [3] Budiyanto A, Widiatmaka W, Sudiono S, Mun'im TWA, Sidhi, Hertian S, Sampurna B, et al. Ilmu kedokteran forensik. Edisi 1. Jakarta: Bagian Kedokteran Forensik Fakultas Kedokteran Universitas Indonesia. 1997
- [4] James JP, Jones R, Karch SB, Manlove J. Principles of Forensic Science. In: Simpson's Forensic Medicine. 13th Ed. London: Hodder & Stoughton. 2011.
- [5] J. B. Reece, L. A. Urry, M. L. Cain, S. A. Wasserman, R. B. Jackson, et al, "Campbell BIOLOGY," San Francisco: Pearson Education, Inc., 2011.
- [6] [https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-\(2018\).pdf](https://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-(2018).pdf)

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 2020



Ade Surya Handika
13518007