Penggunaan *Pattern Matching* untuk Kebutuhan Digital Forensik

Kelvin Kristian/13516101
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung(ITB)
Bandung, Indonesia
13516101@std.stei.itb.ac.id, kelvinkristian4@gmail.com

***Abstract*—Setiap tindakan kejahatan yang ada pasti selalu dilakukan tindakan forensik untuk menelusuri bagaimana kejadian suatu tindak kejahatan. Ilmu sains forensik akan menguraikan seluruh masalah tersebut sehingga dapat ditemukan bukti-bukti yang kuat dan otentik untuk menjatuhkan hukuman pada seorang tersangka. Ilmu sains fisik terdiri dari banyak bidang. Seluruh bidang ini saling bekerja sama untuk mengungkap suatu barang bukti. Salah satu bidang yang memiliki peranan cukup penting di era digital saat ini adalah forensik di bidang digital atau biasa disebut digital forensik. Digital forensik biasanya melakukan pemeriksaan barang bukti pada perangkat-perangkat penyimpanan data elektronik seperti *flash drive, email, harddisk, handphone, pesan singkat* dan masih banyak lagi hal lainnya. Salah satu kegiatan yang paling umum adalah memeriksa konten suatu *email*. Proses pemeriksaan ini bisa menggunakan algoritma *pattern matching* dan nantinya akan memberitahu kita apakah terdapat kata yang kita curigai sebagai barang bukti kejahatan pada *email* tersebut.**

Kata kunci----pola, cocok, Python, forensik, algoritma.

I. PENDAHULUAN

Ilmu forensik adalah ilmu yang mulai dikembangkan pada saat pertama kali saat sidik jari mulai digunakan dalam kegiatan bisnis oleh bangsa cina. Pada saat itu Sir Francis Galton membangun sebuah system untuk mengklasifikasikan sidik jari. Kemudian Sir Edward Henry membangun sistemnya sendiri untuk kepolisian London dan kemudian sistem ini menjadi sistem standar untuk pengidentifikasian sidik jari para kriminal. Setelah itu metode-metode forensik makin berkembang seiring berjalannya waktu dan menjadi semakin canggih.

Gambar 1 Sidik jari untuk dilakukan pemeriksaan forensik

Sumber: https://fit.labs.telkomuniversity.ac.id/memahami-istilah-dari-digital-forensik/

Melihat perkembangannya di zaman dulu, forensik sudah memiliki banyak jenis akan tetapi. Pada zaman sekarang kebanyakan orang menganggap bahwa forensik hanya sebatas kegiatan penyelidikan pada sebuah tindakan kejahatan. Kasus yang biasanya diselidiki adalah kasus seperti penyelidikan pada kasus pembunuhan yang membutuhkan tindakan autopsi untuk dapat mengetahui kronologis suatu pembunuhan beserta motifnya. Namun kejahatan yang ada di dunia ini tidak hanya pembunuhan akan tapi masih banyak lagi jenis kejahatan lainnya. Kejahatan non-konvensional merupakan salah satu jenis kejahatan yang dapat diselidiki kebenarannya melalui forensik. Kejahtan non-konvensional tersebut termasuk kejahatan siber. Kejahatan siber adalah kejahatan yang terjadi melalui atau menggunakan perangkat elektronik. Biasanya hal yang diselidiki dalam kejahatan siber adalah pesan singkat melalui media sosial atau aplikasi pesan singkat. Selain itu bisa juga data-data digital yang lain seperti dokumen penting, sidik jari, rekaman suara atau rekaman video.

Dalam makalah ini, penulis akan membuat suatu program untuk menyelidiki apakah suatu pesan singkat yang merupakan barang bukti kejahatan, mengandung kata-kata yang dicurigai sebagai sumber kejahatan. Jika pesan singkat mengandung kata-kata tersebut maka barang bukti tersebut akan menjadi barang bukti yang kuat dalam persidangan untuk menjatuhkan hukuman pada penjahat. Program yang akan penulis buat adalah program *pattern matching* yang akan menggunakan metode Boyer-Moore yang dikembangkan lebih lanjut di mana metode yang digunakan adalah menggunakan *bad match table*.

II. DASAR TEORI

A. *Forensik*

Forensik dalam bahasa Latin berarti diskusi publik atau debat. Dalam konteks yang lebih modern, forensik digunakan oleh pengadilan atau sistem yudisial. Jika digabungkan dengan kata sains maka sains forensik berarti proses memecahkan kasus kriminal dengan menggunakan metode-metode *scientific*. Bidang yang digunakan dalam sains forensik adalah bidang-

bidang yang menyangkut fisika, kimia dan biologi yang akan berfokus pada pengenalan, pengidentifikasian dan evaluasi dari bukti fisik. Sains forensik sudah menjadi bagian yang penting dalam sistem pengadilan di mana sains forensik menggunakan spektrum sains yang luas untuk mendapatkan informasi yang relevan untuk masalah kriminal dan bukti legal.

Sains Forensik akan membuktikan adanya suatu kejahatan, membuktikan seorang pelaku kejahatan dan bisa juga untuk membuktikan hubungan antara sesuatu dengan tindakan kriminal dengan cara:

1. Memeriksa bukti fisik
2. Tes administrasi
3. Interpretasi data
4. Meluruskan dan meringkas laporan
5. Kesaksian dari ilmuan forensik

Sains forensik sudah menjadi bagian yang penting dalam berbagai kasus kriminal dengan bantuan fakta-fakta yang objektif dan melalui didapat melalui ilmu pengetahuan. Fakta tersebut dapat digunakan untuk membantu argumen pembelaan maupun penuntutan. Fakta-fakta tersebut dapat menjadi bukti yang dapat dipercaya dalam berbagai kasus kriminal.

Gambar 2. Pemeriksaan forensik pada tulang dengan analisis kimia

Sumber: http://blogs.icrc.org/indonesia/ilmu-forensik-2/

Sains forensik melakukan analisis fisik dan kimia terhadap barang bukti yang didapat oleh para penegak hukum dan investigator pada tempat kejadian perkara. Para ahli menggunakan teknik mikroskop, imstrumen yang kompleks, prinsip matematika, prinsip sains, dan literatur referensi untuk menganalisa kelas dan karakteristik dari barang bukti tersebut.

B. *Digital Forensik*

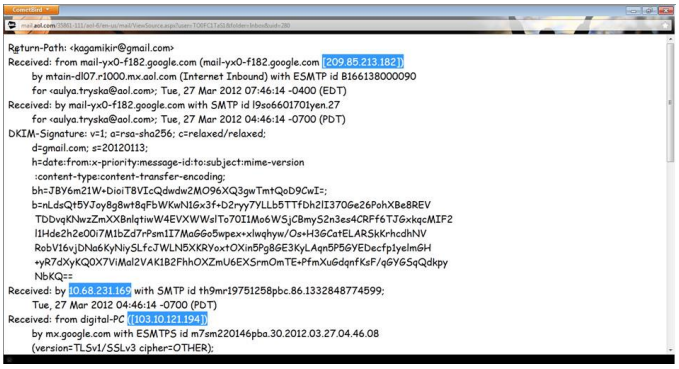
Digital forensik adalah sains forensik yang mengidentifikasi, menjaga, memulihkan, menganalisis, mempresentasikan fakta tentang barang bukti digital yang ditemukan pada komputer atau media penyimpanan digital.

Pemrosesan seluruh digital forensik dimulai dari pengidentifkasian. Hal yang paling penting adalah pengidentifikasian lokasi di mana data disimpan. Pada era digital forensik, kebanyakan data ditemukan pada *hard drives computer, server, flash drive* dan perlengkapan *network.*

Menjaga barang bukti digital forensik merupakan proses yang paling penting dalam proses digital forensik. Integritas suatu data sangat penting dijaga. Jika nilai kebenaran dari data sudah hilang maka data tersebut menjadi tidak berguna lagi di depan pengadilan karena tidak bisa membuktikan apapun.

Pemulihan barang bukti adalah proses yang tidak kalah penting. Pemulihan ada beberapa jenis yaitu seperti pemulihan data yang hilang yang disebabkan prosedur standar dari sistem operasi, atau data yang sengaja dihapus, atau kata sandi dari *file* yang di proteksi, bahkan pada file yang rusak atau korup.

Analisis adalah proses yang akan selalu dilakukan pada setiap investigasi forensik sains. Analisis adalah kegiatan yang menganalisis seluruh data hasil digital forensik. Seluruh data diteliti kemudian dicari nilai kebenarannya yang dihubungkan dengan kasus kriminal yang terjadi.



Gambar 3. Pencarian IP Address pada sebuah log
Sumber: https://aulyatryska.wordpress.com/2012/04/26/uts-itforensic/

Terakhir adalah tahap presentasi di mana setelah hasil pemeriksaan selesai maka hasilnya akan dilaporkan di pengadilan. Semua hasil pemeriksaan yang akan dipresentasikan dibungkus menjadi beberapa dokumen yang sudah rapih. Selanjutnya hasil persidangan tidak lagi ditentukan oleh penyelidik melainkan oleh barang bukti yang sudah disiapkan.

C. *Pattern Matching*

Pattern Matching atau pencocokan pola adalah sebuah ilmu di dalam sains komputer untuk memeriksa dan mencari lokasi sederet data spesifik berdasarkan sebuah pola dari sederet data yang tengah ditelusuri. Pencocokan pola ini harus bersifat *exact matching* atau harus tepat seluruhnya untuk dikatakan cocok. Biasanya pattern matching ini dilakukan saat melakukan fungsi pencarian *find & replace* atau saat menggunakan mesin pencarian seperti *Google search.*

Gambar 4. Pattern Matching

Sumber: http://www.codingalpha.com/naive-pattern-matching-algorithm-c-program/

D. Algoritma Knoth-Morris-Pratt

Algoritma Knuth-Morris-Pratt adalah salah satu algoritma pencarian *string* yang dikembangkan secara terpisah oleh Donald E. Knuth pada tahun 1967 dan James H. Morris bersama Vaughan R. Pratt pada tahun 1966. Algoritma ini akan bekerja lebih efisien dibanbing algoritma *Brute Force* karena algoritma ini lebih menghemat perbandingan jika dibandingkan dengan algoritma *Brute Force* yang membandingkan semuanya satu persatu. Perhitungan penggeseran pada algoritma ini adalah bila terjadi ketidakcocokan pada saat pattern sejajar dengan S[i..i + n -1], bisa dianggap ketidakcocokan pertama terjadi di antara S[i+j] dan P [j], dengan 0<j<n. Berarti, S[i...i + j-1] = P[0..j-1] dan a = S[i+j] tidak sama dengan b = P[j]. Ketika menggeser, perlu diperhatikan bahwa ada kemungkinan sebuah u dari *pattern* yang akan sama dengan sebagian akhir u dari sebagian *String*. Sehingga *pattern* bisa digeser agar awalan u tersebut sejajar dengan akhiran u dengan kata lain pencocokan string akan berjalan lebih efisien bila terdapat table yang dapat menentukan seberapa panjang pattern harus digeser seandainya terdeteksi ketidakcocokan di karakter ke j dari *pattern*. Tabel juga harus memuat next[j] yang merupakan posisi karakter P[j] setelah digeser, sehingga pattern dapat digeser sebesar j-next[j] relatif terhadap *String*. Kompleksitas algoritma KMP adalah O(n+m)

III. Penerapan dan Analisis Algoritma KMP Sebagai Algoritma Pattern Matching Pada Digital Forensik

A. Penerapan Sains Forensik Pada Data Digital

Pada data digital seperti *email* atau pesan singkat biasanya para investigator forensik sains mencari beberapa kata yang diperlukan untuk mengidikasikan bahwa data digital tersebut merupakan barang bukti dari sebuah tindakan kriminal. Misalkan kita hendak mencari kata yang menentukan bahwa seseorang menjual narkoba. Kita hanya perlu menggunakan kata kunci narkoba atau menjual untuk mencari apakah orang tersebut menjual narkoba atau tidak.

Contoh Email: “Besok pukul 4 sore saya akan membawakan narkoba jenis ekstasi di depan hotel ABC di jalan Pantai Utara. Jangan lupa untuk membawa uang yang siudah disepakati yaitu sebesar 100 juta rupiah untuk 5 kilogram ekstasi”

Misalkan menggunakan masukan kata kunci: “narkoba”. Maka hasil akan bernilai *true* karena kata narkoba terdapat pada email tersebut.

B. Penerapan Pattern Matching Menggunakan Algoritma KMP

Pertama-tama *string* harus disiapkan terlebih dahulu kemudian *string* tersebut akan diolah yang dalam kasus digital forensik akan disiapkan sebuah *email* yang dicurigai sebagai barang bukti tindak kejahatan.

Misalkan string yang digunakan adalah ABAABXABAABA. Lalu pattern yang digunakan adalah ABAABA. Pertama-tama kita harus menentukan *border function* dari *pattern*-nya terlebih dahulu.

Border functionnya adalah sebagai berikut:

J	0	1	2	3	4	5
P[j]	A	B	A	A	B	A
K	-	0	1	2	3	4
B[k]	-	0	1	1	2	3

Kemudian proses yang terjadi adalah

1	2	3	4	5	6	7	8	9	10	11	12
A	B	A	A	B	X	A	B	A	A	B	A
A	B	A	A	B	A						
					mismatch						

Karena pada *Pattern* sebelum karakter *Pattern*[6] memiliki kesamaan *prefix* dan sufixnya yaitu AB maka pengecekan tidak perlu dilakukan pada dari awal lagi tetapi dapat dimulai setelah AB yaitu *Pattern*[3].

1	2	3	4	5	6	7	8	9	10	11	12
A	B	A	A	B	X	A	B	A	A	B	A
			A	B	A	A	B	A			
					mismatch						

Karena pada *String*[6] terjadi *mismatch* maka pengecekan dilakukan dari awal karena prefix dan suffix sebelum *Pattern*[3] unik, tidak memiliki kesamaan.

1	2	3	4	5	6	7	8	9	10	11	12
A	B	A	A	B	X	A	B	A	A	B	A
					A	B	A	A	B	A	
					mismatch						

Pengecekan mengalami ketidakcocokan maka dari itu pattern digeser lagi ke kanan sebanyak 1 indeks.

1	2	3	4	5	6	7	8	9	10	11	12
A	B	A	A	B	X	A	B	A	A	B	A
						A	B	A	A	B	A
											match

Setelah mengalami pergeseran maka *String* sudah cocok dengan patternnya maka pattern yang dicari pada *string* dinyatakan sudah ditemukan.

C. Algoritma KMP

Pertama-tama algoritma KMP akan membuat sebuah list yang menghasilkan tabel border function. Tabel tersebut di generate menggunakan fungsi `computeFail()`. Berikut penerapan fungsinya.

```
function computeFail(pattern: string) →
array of integer
(Fungsi akan menerima input pattern
berupa string)
(Kemudian akan dibuat tabel border
function berdasarkan pattern tersebut)

KAMUS
    tab : array of integer
    i   : integer
    j   : integer

ALGORITMA
    Tab[0]  ←  0
    i       ←  1
    j       ←  0

    while (i < len(pattern)) do
            if (pattern[j] ==
            pattern[i]) then
            tab[i] ← (j+1)
            j ← j + 1
            i ← i + 1
        else if (j > 0) then
            j ← tab[j-1]
        else
            tab[i]    0
            i ← i + 1

    return tab
```

Pertama-tama fungsi akan menerima masukan berupa *pattern* bertipe *string*. Kemudian dibuat sebuah *array* untuk menampung nilai *border function*-nya. Langkah berikutnya adalah melakukan iterasi untuk memeriksa karakter demi karakter pada pattern tadi kemudian memberi nilai batasnya satu per satu. Kemudian tabel *border function* tadi dikembalikan.

Lalu fungsi yang di gunakan berikutnya adalah fungsi `kmpMatch()`. Fungsi ini berfungsi untuk melakukan pencocokan dengan menggunakan *border function* yang sudah dibangkitkan.

```
function kmpMatch(text: string,
pattern: string) → array of integer
    found ← array of integer
    b     ← computeFail(pattern)
    i     ← 0
    j     ← 0

    if (len(pattern) == 0) then
        return found

    while (i<len(text)) do
      if (pattern[j] == text[i])
      then
            if (j == len(pattern)-
            1) then
                found.append(i-
                len(pattern)+1)
                if (len(pattern) ==
                1) then
                    i  ←  i + 1
                else:
                    j  ←  b[j-1]
            else:
                i  ←  i + 1
                j  ←  j + 1
       else if (j > 0):
            j  ←  b[j-1]
       else:
            i  ←  i + 1

    return found
```

Pada fungsi `kmpMatch()`, fungsi akan menerima berupa *pattern* dan *text* yang keduanya bertipe *string*. Kemudian akan digunakan sebuah *array of integer* untuk menampung indeks di mana saja suatu kata bersesuaian dengan patternnya. Border function di tamping kedalam variable b dengan cara memanggil method `computeFail()`. Jika terdapat sebuah string yang cocok dengan patternnya maka akan dicatat di indeks mana pertama kali terjadi kecocokan pada string.

D. Pencarian Suatu Kata pada Email yang Digunakan dalam Digital Forensik

Misalkan kita mendapatkan beberapa buah email yang berasal dari tersangka transaksi penjual narkoba dan kita hendak memeriksa apakah email tersebut apakah benar-benar email yang berisikan sesuatu mengenai narkoba. Maka dari itu kita dapat mengiterasi semua email yang kita miliki lalu dicocokan dengan kata kunci yang kita harapkan berada pada email tersebut.

```
D:\Academic\Institute ITB\Jurusan\Informatika\Semester 4\I
Please input the pattern     : narkoba
377
Nanti malam jam 4 sore kita bertemu di samping dermaga.
Barangnya sudah saya siapkan, impor dari America yaitu
narkoba berupa ekstasi murni sebanyak 100 kg.
Harga berdasarkan harga yang sudah di sepakati yaitu
2.000.000.000 . Jangan sampai ada yang mengikuti anda,
sampai anda membawa siapapun atau bahkan membawa polisi
maka perjanjian akan batal dan kamu akan saya butuh.
This is the border function  : [0, 0, 0, 0, 0, 0, 0]
Index the pattern found      : [113]
```

Pada pemeriksaan email tersebut pattern narkoba dibuat *border function*-nya dan menghasilkan *border function* yang menghasilkan 0 semua. Hasil tersebut terjadi karena *pattern* narkoba tidak memiliki *subset prefix* dan *suffix* yang sama sehingga tidak memungkinkan untuk untuk melakukan penelusuran *pattern* yang tidak dilakukan dari depan.

Kemudian *pattern* digunakan untuk melakukan *matching* pada seluruh kalimat di dalam *email*. *Matching* dimulai dari kata pertama kemudian terus dilanjutkan sampai akhir dari *email*. Setelah dilakukan penelusuran. Ditemukan bahwa *pattern* narkoba ditemukan pada *index* atau karakter ke 113. Ditemukannya pola narkoba yang dicari pada *email* tersebut mengindikasikan bahwa *email* tersebut adalah barang bukti terjadinya tindakan kejahatan. Maka dari itu *email* tersebut dapat dijadikan barang bukti dalam pengadilan.

111	112	113	114	115	116	117	118	119
U		N	A	R	K	O	B	A
N	A	R	K	O	B	A		
miss								

111	112	113	114	115	116	117	118	119
U		N	A	R	K	O	B	A
	N	A	R	K	O	B	A	
	miss							

111	112	113	114	115	116	117	118	119
U		N	A	R	K	O	B	A
		N	A	R	K	O	B	A
								match

Kemudian kita akan mengambil contoh pada email lain.

```
D:\Academic\Institute ITB\Jurusan\Informatika\Semester 4\I
Please input the pattern     : narkoba
315
Terima kasih karena telah melakukan pembelian ekstasi
kemarin ini. Hari ini pesanan narkoba jenis kokain milik
anda sudah dapat diambil pada agen kami. Anda dapat
mengambilnya di samping hotel Alexis, Jakarta pada pukul
5 sore ini. Harga seperti yang sudah disepakati kemarin
untuk 10kg kokain senilai 500.000.000
This is the border function  : [0, 0, 0, 0, 0, 0, 0]
Index the pattern found      : [84]
```

Pada email kedua ini ternyata ditemukan juga kata narkoba pada email tersebut. Pertama kita akan membangkitkan *border function* terlebih dahulu. Bisa dilihat karena menggunakan kata yang sama seperti sebelumnya yaitu narkoba maka border function yang dihasilkan pun juga 0 semua di mana tidak terdapat subset prefix dan suffix yang sama pada pattern tersebut.

Kemudian pattern juga digunakan untuk memeriksa email tersebut. Setelah dilakukan pemeriksaan maka email tersebut mengandung kata narkoba di mana artinya email tersebut bisa dijadikan barang bukti yang menunjukan terjadinya tindakan kriminal berupa terjadinya transaksi narkoba. Maka dari itu dapat disimpulkan bahwa tersangka yang sedang diperiksa kejahatannya dapat dijatuhi hukuman yang cukup berat karena memiliki 2 buah email yang mengandung unsur kejahatan yaitu transaksi narkoba.

IV. KESIMPULAN

Forensik adalah kegiatan melakukan pemeriksaan atau penyelidikan terhadap suatu tindakan kejahatan. Salah satu cabang dari forensik adalah digital forensik di mana pemeriksaan terjadi benda-benda digital seperti data-data pada email, pesan singkat, harddisk, flashdrive dan lainnya. Dalam membantu kinerja para pemeriksa, dapat digunakan algoritma *pattern matching* KMP yaitu algoritma yang mencocokan sebuah *pattern* pada *string* untuk memeriksa apakah ada indikasi kejahatan di dalam string tersebut. Tentu saja *keyword* atau *pattern* yang digunakan pasti kata-kata yang mengandung unsur kejahatan. Cara pencocokannya adalah dengan menggunakan *border function* untuk mempercepat pencocokan per karakternya.

a.

UCAPAN TERIMA KASIH

Pertama, penulis ingin mengucapkan puji dan syukur kepada Tuhan Yang Maha Esa atas rahmat-Nya sehingga makalah ini dapat terselesaikan dengan baik. Saya juga mengucapkan terima kasih kepada Ibu Nur Ulfa Maulidevi selaku dosen pengajar IF2211 Strategi Algoritma yang sudah mengajarkan serta menjadi sumber inspirasi dalam penulisan makalah ini

REFERENCES

[1] https://science.howstuffworks.com/forensic-lab-technique1.htm

[2] https://penakuliah.wordpress.com/2015/12/01/digital-forensic-atau-forensik-digital/, diakses pada 12 Mei 2018

[3] http://www.ijecs.in/issue/v3-i6/20%20ijecs.pdf, diakses pada 12 Mei 2018

[4] https://www.crimesceneinvestigatoredu.org/what-is-forensic-science/, diakses pada 12 Mei 2018

[5] https://interworks.com/blog/bstephens/2016/02/05/what-digital-forensics/, diakses pada 12 Mei 2018

[6] https://www.technopedia.com/definition/8801/pattern-matching, diakses pada 12 Mei 2018

[7] https://towagapatma.wordpress.com/2014/02/15/tentang-algoritma-knuth-morris-pratt/, diakses pada 12 Mei 2018

[8] http://informatika.stei.itb.ac.id/~rinaldi.munir/Stmik/2017-2018/Pencocokan-String-(2018).pdf, diakses pada 12 Mei 2018

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, 6 April 2018

Kelvin Kristian
13516101