

Analisis Pengukuran Perubahan *Website* Menggunakan Algoritme Jaro-Winkler *Distance* Sebagai Acuan Pendeteksi *Website Defacement*

Nurdan Darajat Ulya M.¹⁾, Teguh Bharata Adji²⁾, Dani Adhipta³⁾

Jurusan Teknik Elektro dan Teknologi Informasi

Jl.Grafika No.2 , Fakultas Teknik

Universitas Gadjah Mada

aka_nurdan@ugm.ac.id ¹⁾, adji@ugm.ac.id ²⁾, dani@ugm.ac.id ²⁾

Abstract- *The Popularity of web-based services enables many companies and organizations to compete by using website as publication media. On the other hand, recently website content become one of primary targets of cyber world's attack in the form of website defacement. Due to the seriousness of the impact of these attacks and the long time response and recovery attempt, demanding for engineers and researchers to develop an early detection system against website defacement. This study proposed a new initiative in the form of method to detect website defacement by applying string metric Jaro-Winkler distance algorithms as an approach for measuring the rate of change of 25,500 records from top 10 Indonesian universities websites at Webometric and 5 highest-ranked Indonesian-language news website by Alexa.com observed during the two and a half months. This research has analyzed the pattern of changes in the website and specify a valid threshold value as a reference for building a website defacement detection system and has managed to detect more than 93% of the testing of the whole object of research with a warning system as expected.*

Keyword— *website defacement; change measurement; defacement detection; jaro-winkler distance*

Intisari— Popularitas layanan berbasis *web* membuat sebagian besar perusahaan dan organisasi mampu bersaing menggunakan *web* sebagai sarana publikasi mereka. Di sisi lain, hari ini konten *website* menjadi salah satu target utama serangan di dunia maya dalam bentuk *website defacement*. Mengingat seriusnya dampak yang ditimbulkan dari serangan jenis ini, serta lamanya waktu reaksi dan pemulihan pasca kejadian, menuntut keseriusan bagi teknisi dan peneliti untuk mengembangkan sistem deteksi dini terhadap *website defacement*. Penelitian ini mengusulkan inisiatif baru berupa metode pendeteksian *website defacement* dengan menerapkan algoritme *string metric* Jaro-Winkler *distance* sebagai pendekatan untuk pengukur tingkat perubahan suatu *website* terhadap 25.500 *record* dari 10 *website* universitas di Indonesia pemilik peringkat tertinggi versi Webometric dan 5 *website* berita berbahasa Indonesia peringkat tertinggi versi Alexa.com hasil observasi selama dua setengah bulan. Penelitian ini berhasil menganalisis pola perubahan *website* dan menentukan nilai *threshold* yang *valid* sebagai acuan untuk membangun *prototype* sistem pendeteksi *website defacement* dan telah berhasil mendeteksi lebih dari 93% dari pengujian terhadap seluruh objek penelitian dengan sistem peringatan seperti yang diharapkan.

I. PENDAHULUAN

Banyaknya perusahaan dan organisasi yang menggunakan layanan *web* sebagai sarana untuk mempromosikan kegiatan bisnis mereka mengakibatkan layanan *internet dan web* telah menjadi infrastruktur utama untuk kegiatan bisnis saat ini [1], Sehingga dalam arsitektur jaringan, *web server* umumnya diposisikan pada ujung tombak suatu organisasi. Hal ini menjadi salah satu penyebab mengapa *web server* menjadi salah satu target utama serangan pelaku *defacement* atau yang sering disebut dengan istilah *defacer*. Mengacu pada halaman Zone-H.org [2] yang merupakan *website* acuan utama para *defacer* dalam mempublikasikan daftar halaman *website* yang telah mereka rusak, jumlah halaman yang berhasil dirusak meningkat pesat setiap tahunnya. Tercatat sejak tahun 2010 sampai dengan 2013, rata-rata setiap tahun lebih dari 1,2 juta halaman *website* berhasil ter-*deface* [3]. Gangguan layanan *web* jenis ini dapat menyebabkan banyak masalah, seperti berkurangnya jumlah pengakses, penurunan kredibilitas perusahaan, dan sejenisnya [1].

Mempertimbangkan sifat dinamis konten *website* yang secara normal terus berubah sesuai *update* yang dilakukan oleh pengelola *web*, dan cenderung dilakukan pada level *database*, maka diperlukan metode pendeteksian *defacement* halaman *web* dengan pendekatan tingkat kemiripan atau perbedaan halaman *website* dalam rentang waktu tertentu berbasis nilai ambang (*threshold*) sebagai metrik pendeteksian *defacement* suatu halaman *web* yang bersifat dinamis.

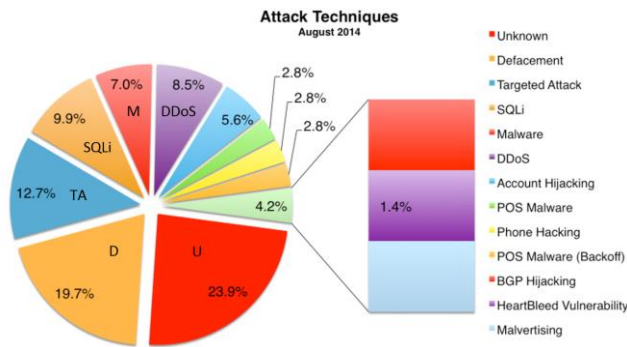
II. TINJAUAN PUSTAKA DAN LANDASAN TEORI

A. Tinjauan Pustaka

1. Trend Serangan di dunia *Cyber*

Website defacement merupakan salah satu jenis serangan yang saat ini paling sering terjadi di dunia *cyber*, sekitar 19,7 % dari total serangan *cyber* saat ini berupa tindakan *Website Defacement*, 12,7 % berjenis *Targeted Attack* yaitu jenis serangan yang secara khusus ditargetkan pada organisasi atau perusahaan tertentu, 9,9% berupa serangan *SQL Injection*, 7% berupa serangan *Malware*, 8,5% berupa *DDoS*, 5,6% berupa

serangan *Account Hijacking*, dan sisanya beberapa serangan lain yang belum diketahui jenisnya [4]. (Gambar 2.2)



Gambar 2.1 Pie Chart perbandingan teknik cyber attack [4]

2. Webometric

Webometric adalah salah satu perangkat atau sistem untuk mengukur atau memberikan penilaian terhadap kemajuan seluruh universitas atau perguruan tinggi terbaik di dunia (*World Class University*) melalui *website* universitas tersebut [8]. Peringkat Webometrics yang dimulai sejak tahun 2004 ini bisa menjadi acuan bagi calon mahasiswa dalam memilih universitas terbaik yang ada di Indonesia maupun di dunia. Lembaga ini berafiliasi dengan Dewan Riset Nasional Spanyol. Dalam setahun, badan ini mengeluarkan daftar peringkat, yaitu Bulan Januari dan Juli [9].

3. Alexa

Alexa.com adalah *website* yang mengklaim dirinya sebagai *website* yang secara eksklusif fokus memberikan sarana analisis terkaya dan paling berarti bagi para pelanggannya. Alexa merupakan pelopor global dalam dunia wawasan analitis, merupakan pengembang layanan analisis *web* yang paling kuat dan akurat diantara penyedia lainnya. Estimasi yang dilakukan Alexa didasarkan pada panel lalu lintas global yang mewakili jutaan pengguna internet yang menggunakan salah satu dari lebih dari 25.000 jenis browser yang berbeda [10].

Website berjenis konten berita merupakan salah satu jenis *website* yang paling banyak diakses selain *website-website* hiburan dan *website* blog, salah satu *website* berita lokal yang masuk 10 besar situs terpopuler di Indonesia adalah situs berita detik.com yang bahkan mengalahkan peringkat *website* terkenal penyedia *free blog hosting* wordpress.com. Inilah salah satu alasan digunakannya *website* berita sebagai salah satu sarana objek dalam penelitian ini.

4. Performa Algoritme Jaro-Winkler Distance

Dalam penelitian [5] telah dilakukan perbandingan performa berbagai jenis algoritme *string metric* dalam mengukur kemiripan *string*, yaitu berupa perbandingan nilai *F-measure* masing-masing algoritme *string metric* dalam menangani *string* berupa nama orang yang terdiri dari *given names*, *surnames*, dan *full names* (Tabel 2.1)

Tabel 2.1 Rata-rata nilai *F-measure* (hasil terbaik ditampilkan dengan angka tebal, hasil terburuk ditampilkan dengan garis bawah) [5:12]

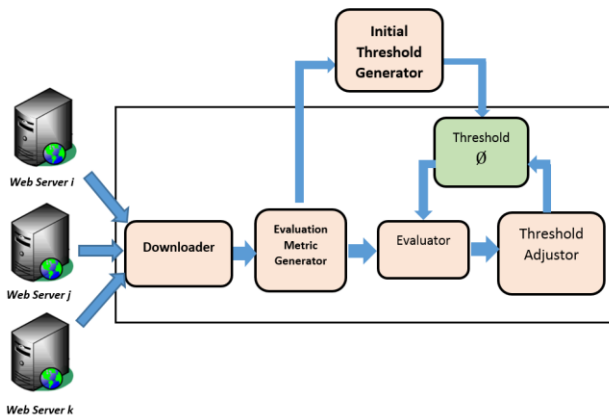
	Midwives			COMPLETE surnames
	given names	sur- names	full names	
Soundex	.342	.341	.376	.485
Phonex	.423	.369	.499	.579
Phonix	.339	.330	.368	.617
NYSIIS	<u>.275</u>	<u>.296</u>	<u>.299</u>	<u>.351</u>
DMetaphone	.304	.306	.330	.410
FuzSoundex	.327	.311	.359	.396
Leven dist	.658	.513	.737	.624
Dam-L dist	.659	.517	.739	.625
Bag dist	.597	.522	.670	.616
SWater dist	.889	.579	.802	.617
LCS-2	.915	.564	.877	.514
LCS-3	.909	.529	.866	.500
1-grams	.839	.588	.787	.627
2-grams	.885	.498	.867	.519
3-grams	.783	.442	.833	<u>.416</u>
Pos 1-grams	.890	.574	.724	.653
Pos 2-grams	.880	.473	.697	.508
Pos 3-grams	.768	<u>.416</u>	.659	<u>.416</u>
Skip grams	.844	.496	.825	.521
Compr BZ2	<u>.458</u>	.547	<u>.568</u>	.633
Compr ZLib	.532	.456	.684	.481
Jaro	.853	.601	.829	.712
Winkler	.891	.588	.868	.707
SortWink	.803	.580	.809	.707
PermWink	.888	.598	.883	.707
Editex	.631	.561	.706	.646
SAPS dist	.656	.426	.710	.532

Pada tabel di atas Jaro *distance metric* memiliki nilai *F-measure* terbaik sebesar 0,601 untuk kategori *surenames* dan 0,712 untuk kategori *complete surenames*. Penelitian [6] menerapkan Gabungan algoritme *string comparator* Jaro *distance* dan Winkler *distance* sebagai sistem pendeteksian plagiarisme pada dokumen berbahasa Indonesia. Dalam ujicobanya, aplikasi dapat berjalan dengan kinerja yang sangat baik, begitu juga dengan penelitian [7], performa yang lebih baik ditunjukkan oleh algoritme Jaro-Winkler *distance*, dalam hal waktu eksekusi dibandingkan dengan algoritme *hybrid comparator edit distace*

B. Landasan Teori

1. Proses Pendeteksian Website Defacement

Semua *tools* pendeteksi *website defacement*, kebanyakan memiliki konsep yang sama, yaitu membuat salinan master data acuan yang terpercaya dari halaman *web* yang akan dipantau, kemudian disimpan di lokasi yang aman, selanjutnya konten data halaman web terkini dibandingkan dengan master data acuan yang terpercaya dengan menambahkan fitur notifikasi jika muncul adanya ketidakcocokan. Perbandingan dapat dilakukan secara berkala, berdasarkan skema pemantauan pasif, atau dengan pengiriman *source code* dari *web* kepada klien secara aktif. Contoh skema pendeteksi *website defacement* secara *remote* ditunjukkan pada Gambar 2.2



Gambar 2.2 Proses Pendeteksian Web Defacement secara jarak jauh [1].

2. Jaro-Winkler Distance

Jaro-Winkler distance merupakan varian dari Jaro distance metric yang dikembangkan kembali oleh William E. Winkler, yaitu sebuah algoritme untuk mengukur kesamaan antara dua string [6], biasanya algoritme ini digunakan di dalam pendeteksian duplikat. Semakin tinggi Jaro-Winkler distance untuk dua string, semakin tinggi kemiripan kedua string tersebut. Skor normalnya yaitu antara 0 dan 1, nilai 0 menandakan tidak ada kesamaan, dan 1 adalah sama persis [7]. Algoritme Jaro-Winkler distance memiliki kompleksitas waktu quadratic runtime complexity yang sangat efektif dan dapat bekerja lebih cepat dari algoritme edit distance. Dasar dari algoritme ini memiliki tiga bagian:

1. Menghitung panjang string.
2. Menemukan jumlah karakter yang sama di dalam dua string, dan
3. Menemukan jumlah transposisi

Pada algoritme dasar Jaro digunakan rumus untuk menghitung jarak (dj) antara dua string yaitu s1 dan s2 adalah :

$$d_j = \frac{1}{3} + \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right) \tag{2.1}$$

Keterangan :

- m = jumlah karakter yang sama persis
- |s₁| = panjang string 1
- |s₂| = panjang string 2
- t = jumlah transposisi (Karakter sama yang tertukar)

Jarak teoritis dua buah karakter yang disamakan dapat dibenarkan jika tidak melebihi nilai :

$$\left(\frac{\max(|s_1|, |s_2|)}{2} \right) - 1$$

Sedangkan algoritme Jaro-Winkler distance menggunakan prefix scale (p) yang memberikan tingkat penilaian yang lebih, dan prefix length (l) yang menyatakan panjang awalan yaitu panjang karakter yang sama dengan string yang dibandingkan sampai ditemukannya ketidaksamaan. Bila string s₁ dan s₂ yang diperbandingkan, maka Jaro-Winkler distance-nya (dw) adalah :

$$d_w = d_j + (lp(1-d_w)) \tag{2.2}$$

dimana :

- d_j = hasil perhitungan kemiripan string s₁ dan s₂ dengan algoritme Jaro distance
- l = adalah panjang karakter yang sama pada awalan string sebelum ditemukan adanya ketidaksamaan dengan batas maksimum sampai 4 karakter.
- p = nilai standar untuk konstanta dalam karya Winkler yaitu p = 0,1

Nilai perhitungan yang didapat dari 0 hingga 1. Dengan nilai 0 setara dengan tidak ada kemiripan dan 1 adalah sama persis.

Contoh Perhitungan Jaro-Winkler distance :

Jika string S₁ adalah “<h1>NURDAN</h1>” dan S₂ adalah “<h1>NURDIN</h1>” maka:

nilai m = 14, S₁ = 15, S₂ = 15 dan tidak ada transposisi atau karakter sama yang tertukar. Maka t = 0.

Maka nilai Jaro distance adalah :

$$d_j = \frac{1}{3} \times \left(\frac{14}{15} + \frac{14}{15} + \frac{14-0}{14} \right)$$

Kemudian bila diperhatikan susunan s₁ dan s₂ dapat diketahui nilai l = 8, sedangkan l max yang diperbolehkan hanyalah 4, sehingga l = 4 dengan nilai konstan p = 0.1. Maka nilai Jaro-Winkler distance adalah

$$d_w = 0.955 + (4 \times 0.1 (1 - 0.955)) = 0.973$$

Jika string s₁ = “ULYA” dan s₂ = “LUYA” maka:

m = 11, s₁ = 11, s₂ = 11, sedangkan nilai t = 1, hal ini dikarenakan terdapat karakter yang sama tapi tertukar urutannya, yaitu karakter U yang tertukar urutannya dengan karakter L. Karakter seperti <, B, >, U, L, Y, A, <, /, B, dan > dianggap dalam urutan yang sama. Maka nilai Jaro distance adalah :

$$d_j = \frac{1}{3} \times \left(\frac{11}{11} + \frac{11}{11} + \frac{11-1}{11} \right) = 0.9697$$

Kemudian bila diperhatikan susunan s₁ dan s₂ dapat diketahui nilai l = 3, dan dengan nilai konstan p = 0.1.

Maka nilai Jaro-Winkler distance adalah:

$$d_w = 0.9697 + (3 \times 0.1 (1 - 0.9697)) = 0.97879$$

3. Preprocessing Data

Sebelum konten website diproses dengan teknik komparasi Jaro-Winkler distance, terlebih dahulu dilakukan preprocessing data melalui beberapa tahap penyaringan konten web yang masih berupa teks source code html website. Adapun proses penyaringan tersebut antara lain dengan :

- Menghilangkan karakter angka dan white space
- Menghilangkan berbagai tanda baca
- Mengkonversi semua huruf ke lower case [6]

4. Initial Threshold

Dalam membangun sistem pendeteksian website defacement, Initial threshold sangat dibutuhkan sebagai acuan untuk mengetahui apakah halaman website yang

bersangkutan terkena defacement ataukah tidak. Jika nilai perubahan suatu halaman *website* melebihi nilai *threshold* yang ditentukan, maka *website* tersebut dianggap mengalami *defacement*. *Initial threshold* pada masing-masing halaman *website* dihasilkan berdasarkan observasi dan kalkulasi nilai perubahan dalam rentang waktu tertentu [1]. Dalam penelitian ini digunakan nilai kemiripan yang terkecil $\min\{dw\}$ dikurangi nilai simpangan baku perubahan kemiripan Jaro-Winkler *distance* sebagai acuan ambang (*threshold*) perubahan halaman *website*

$$\emptyset = \min\{ dw \} - \sigma(\Delta dw) \quad (2.5)$$

Keterangan :

- \emptyset = *Threshold*
- dw = Nilai Jaro-Winkler *distance*
- Δdw = *Delta dw*, selisih nilai kemiripan
- $\sigma(\Delta dw)$ = *Standard* penyimpangan Δdw

III. METODOLOGI

A. Bahan Penelitian

1. Data Sekunder

Data sekunder yang digunakan dalam penelitian ini berupa daftar alamat *website* universitas di Indonesia pemilik peringkat tertinggi versi webometric yang didapatkan lewat situs webometric.info, dan beberapa *website* berita berbahasa Indonesia pemilik ranking tertinggi yang didapatkan melalui situs Alexa. Alasan utama dipilihnya data- dari *website* tersebut karena dianggap dapat mewakili portal situs-situs dinamis di Indonesia dengan tingkat perubahan yang cukup signifikan yang telah terbukti melalui peringkat Webometric dan Alexa *rank*.

2. Data Primer

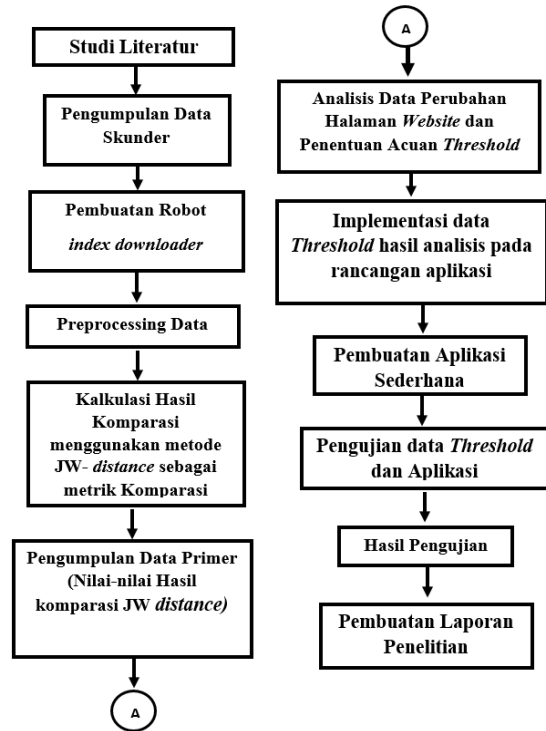
Data primer adalah data yang diperoleh secara langsung melalui metode observasi selama dua setengah bulan sejak 25 Desember 2014 hingga 10 Maret 2015 terhadap data-data sekunder yang telah dipilih. Data yang diambil berupa data waktu pengambilan data, nama atau alamat *website*, dan nilai hasil komparasi similaritas Jaro-Winkler *distance*.

B. Alur Penelitian

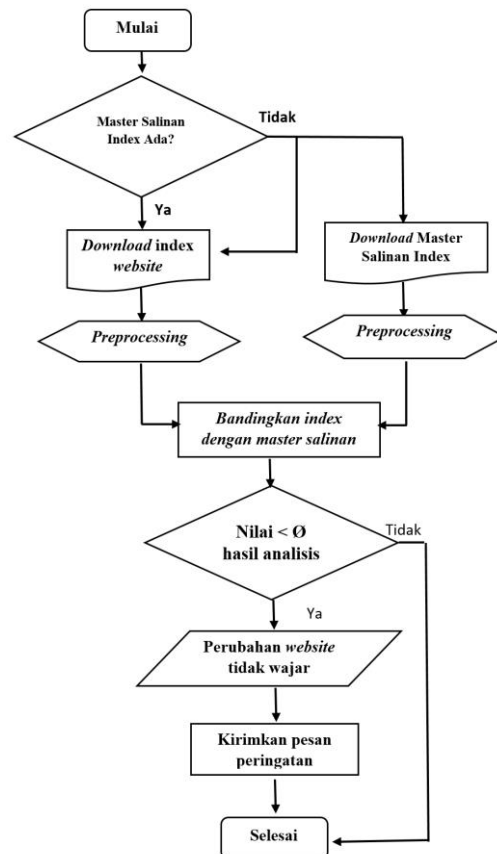
Proses jalannya penelitian dapat dilihat melalui Gambar 3.1.

C. Pembuatan Aplikasi Sederhana

Data-data hasil analisis yang dilakukan pada data-data primer yang didapatkan melalui observasi dan kalkulasi menggunakan metode *string comparison* Jaro-Wikler *distance*,



Gambar 3.1 Diagram Penelitian



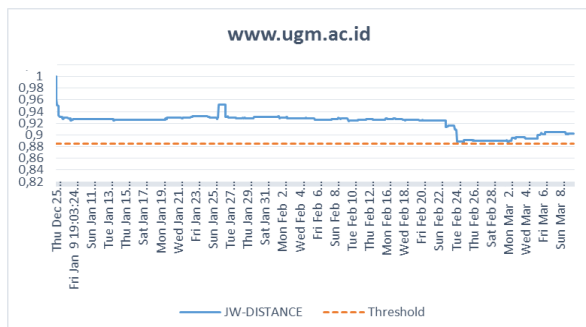
Gambar 3.2 Diagram Alir Sistem Pendeteksi *website defacement*

selanjutnya digunakan sebagai acuan untuk membangun aplikasi sederhana untuk membuktikan bahwa hasil analisis dalam penelitian ini dapat diaplikasikan pada suatu sistem pendeteksi *website defacement*. Pembuatan aplikasi sederhana ini melewati tahap-tahapan sebagai berikut :

1. Rancangan Aplikasi dan Implementasi Hasil Analisis (Gambar 3.2)
2. Instalasi Aplikasi pendukung (PHP, Python dll.)
3. Penulisan *Script* Aplikasi
4. Penulisan *Script* Notifikasi

IV. HASIL ANALISIS DAN PEMBAHASAN

Analisis data primer yang digunakan adalah teori analisis deskriptif, yaitu dengan mencari nilai kemiripan Jaro-Wikler terkecil yaitu $\min\{dw\}$ yang menunjukkan perubahan terbesar yang dialami indeks *website* yang diuji, dan mencari nilai standar penyimpangan (σ) dari delta perubahan nilai kemiripan (Δ_{dw}), kemudian menentukan nilai \emptyset *threshold* yang didapatkan melalui pengurangan nilai kemiripan terkecil $\min\{dw\}$ dengan nilai standar penyimpangan dari delta perubahan nilai kemiripan $\sigma(\Delta_{dw})$. Salah satu contoh grafik pola perubahan nilai kemiripan Jaro-Winkler indeks *website* www.ugm.ac.id dapat dilihat pada Gambar 4.1



Gambar 4.1 Grafik hasil perbandingan kemiripan Jaro-Winkler *distance* beserta *threshold*-nya pada *website* www.ugm.ac.id

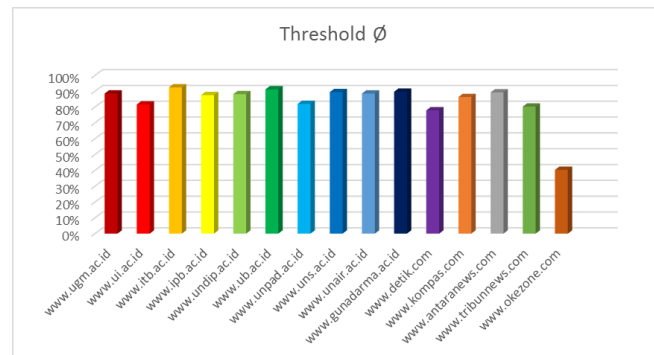
Tabel 4.1 Perbandingan nilai *threshold* masing-masing *website*

No	Alamat Website	JWD Terkecil $\min\{dw\}$ x 100%	STDEV $\sigma(\Delta_{dw})$ x 100%	Threshold \emptyset x 100%
1	www.ugm.ac.id	88,72%	0,161%	88,56%
2	www.ui.ac.id	82,03%	0,438%	81,59%
3	www.itb.ac.id	92,65%	0,214%	92,44%
4	www.ipb.ac.id	88,09%	0,557%	87,54%
5	www.undip.ac.id	88,39%	0,268%	88,12%
6	www.ub.ac.id	91,34%	0,143%	91,19%
7	www.unpad.ac.id	82,72%	0,76%	81,96%
8	www.uns.ac.id	89,6%	0,175%	89,44%
9	www.unair.ac.id	90,07%	0,193%	88,54%
10	www.gunadarma.ac.id	89,87%	0,216%	89,66%
11	www.detik.com	78,51%	0,589%	77,92%
12	www.kompas.com	86,64%	0,297%	86,34%
13	www.antaraneews.com	89,52%	0,245%	89,27%
14	www.tribunnews.com	80,819%	0,161%	80,28%
15	www.okezone.com	42,878%	2,56%	40,31%

A. Rangkuman hasil Analisis terhadap 15 *website*

Hasil analisis terhadap ke-15 *website* menghasilkan data-data seperti pada Tabel 4.1. Dengan data nilai *threshold website* yang menjadi objek penelitian ini maka dapat dibuat aplikasi sederhana sebagai *prototype* sistem pendeteksi dini terhadap serangan *website defacement*

Dari Tabel 4.1 dapat dibuat grafik histogram perbandingan nilai *threshold* masing-masing *website* seperti pada Gambar 4.2.



Gambar 4.2 Grafik Nilai *Threshold* Kemiripan 15 *website* objek penelitian

B. Pengujian Hasil

Tabel 4.2 Hasil Pengujian Nilai Minimal Kemiripan JWD Data *Testing* Dalam Keadaan *Website* Normal Dibandingkan *Threshold* Acuan Dari Data Hasil Observasi

No	Alamat Website	JWD Terkecil $\min\{dw\}$ Data <i>Testing</i> x 100%	Threshold \emptyset x 100%	Valid if $\emptyset < \min\{dw\}$
1	www.ugm.ac.id	91,274%	88,56%	Valid
2	www.ui.ac.id	82,594%	81,59%	Valid
3	www.itb.ac.id	92,65%	92,44%	Valid
4	www.ipb.ac.id	89,19%	87,54%	Valid
5	www.undip.ac.id	89,58%	88,12%	Valid
6	www.ub.ac.id	91,34%	91,19%	Valid
7	www.unpad.ac.id	91,19%	81,96%	Valid
8	www.uns.ac.id	89,42%	89,44%	Valid
9	www.unair.ac.id	90,08%	88,54%	Valid
10	www.gunadarma.ac.id	100%	89,66%	Valid
11	www.detik.com	86,86%	77,92%	Valid
12	www.kompas.com	88,88%	86,34%	Valid
13	www.antaraneews.com	90,18%	89,27%	Valid
14	www.tribunnews.com	85,88%	80,28%	Valid
15	www.okezone.com	76,83%	40,31%	Valid

- Pengujian dengan Data *Testing*

Data *testing* diperoleh dari hasil observasi lanjutan terhadap lima belas *website* objek penelitian selama dua puluh hari sehingga masing-masing *website* menghasilkan *dataset* observasi sebanyak 450 *record* data. Masing-masing *dataset* disaring dengan menghilangkan *record* yang memiliki nilai 0 pada variabel nilai kemiripan Jaro-Winkler *distance*, dengan asumsi nilai 0 berarti kondisi *website* sedang dalam keadaan *down*, atau tidak bisa diakses, sehingga nilai-nilai yang diinginkan adalah nilai kemiripan Jaro-Winkler *distance* dalam keadaan *website* normal dan dapat diakses. Hasil pengujian dapat dilihat pada Tabel 4.2

Tabel 4.3 Hasil Pengujian Nilai Kemiripan JWD Dalam Keadaan Ter-Deface Dibandingkan *Threshold* Acuan dari Data Hasil Observasi

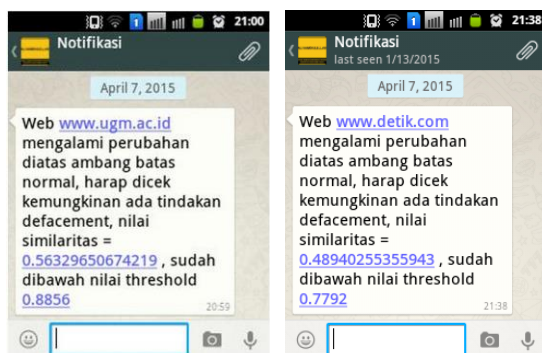
No	Alamat Website	Similaritas JWD Hacker Signature $dw \times 100\%$	<i>Threshold</i> $\emptyset \times 100\%$	Terdeteksi jika $dw < \emptyset$
1	www.ugm.ac.id	56,32%	88,56%	Terdeteksi
2	www.ui.ac.id	50,04%	81,59%	Terdeteksi
3	www.itb.ac.id	54,68%	92,44%	Terdeteksi
4	www.ipb.ac.id	55,58 %	87,54%	Terdeteksi
5	www.undip.ac.id	51,88%	88,12%	Terdeteksi
6	www.ub.ac.id	54,80%	91,19%	Terdeteksi
7	www.unpad.ac.id	61,28%	81,96%	Terdeteksi
8	www.uns.ac.id	52,91%	89,44%	Terdeteksi
9	www.unair.ac.id	52,22%	88,54%	Terdeteksi
10	www.gunadarma.ac.id	57,56%	89,66%	Terdeteksi
11	www.detik.com	48,94%	77,92%	Terdeteksi
12	www.kompas.com	50,77 %	86,34%	Terdeteksi
13	www.antaraneews.com	52,90%	89,27%	Terdeteksi
14	www.tribunnews.com	49,06%	80,28%	Terdeteksi
15	www.okezone.com	49,77%	40,31%	Tidak Terdeteksi

• Pengujian Aplikasi Sederhana

Pengujian yang dilakukan berupa simulasi serangan *defacement* terhadap halaman indeks *website*, pengujian dilakukan menggunakan *virtual address web server* yang mengandung indeks berupa *hacker signature* “BlankJac” dengan *source code* seperti pada halaman *web* <http://pastebin.com/j6sKNdFg>, telah diperoleh data-data pengujian seperti pada Tabel 4.3

C. Hasil Pengujian Notifikasi

Beberapa detik setelah file `/etc/hosts` diubah dengan mengarahkan *resolve* domain *website* ke arah *virtual address web server* maka diperoleh pesan notifikasi melalui ponsel dengan aplikasi WhatsApp *Messenger* seperti Gambar 4.3



Gambar 4.3 Pesan Notifikasi *website* ketika terjadi *defacement*

Secara keseluruhan dari lima belas *website* objek penelitian, sistem pendeteksi *website defacement* mampu memberikan peringatan yang diharapkan terhadap 93,3% objek penelitian yang terdiri dari empat belas *website* yang berhasil dideteksi dan hanya satu *website* saja yang tidak mampu dideteksi

V. KESIMPULAN DAN SARAN

1. Kesimpulan

- Penelitian ini telah menghasilkan nilai *threshold* perubahan halaman *website* sebagai dasar untuk menentukan kriteria *web defacement* dengan tingkat akurasi hingga 93,3% dari seluruh data pengujian.
- Pendeteksian *website defacement* menggunakan acuan *threshold* ini memiliki kelebihan dalam mendeteksi tindakan *web defacement* yang dilakukan dengan menggunakan *hacker signature* yang kebanyakan digunakan oleh para *defacer*.
- Pendeteksian *website defacement* menggunakan acuan *threshold* ini masih memiliki kelemahan pada tindakan *defacement* ringan yang tidak memberikan perubahan cukup signifikan pada *source code* halaman indeks *website* seperti penyusupan *mini script*, *embed web content* seperti Flash, dan Silverlight.

2. Saran

Untuk melengkapi validitas *threshold* yang dihasilkan masih ada pekerjaan penting yang sebaiknya dilakukan pada penelitian mendatang yaitu :

- Melakukan penambahan tahapan *preprocessing* data berupa fungsi *outlier filtering* untuk mengantisipasi adanya anomali perubahan pada *website*, sehingga terdeteksi sebagai *error website* (bukan sebagai *web defacement*).
- Karena objek pengukuran perubahan *website* dalam penelitian ini masih dibatasi hanya terhadap *source code* HTML pada halaman indeks *website*, maka pada penelitian ke depan diharapkan dapat dilanjutkan dengan mengukur juga kemiripan pada konten-konten citra seperti *image*, *embed web content* seperti Flash, Silverlight dan sejenisnya.

DAFTAR PUSTAKA

- [1] Kim Woonyon, Lee Jeongmoon, Park Eungki, Sangwook Kim “Advanced Mechanism for Reducing False Alarm Rate in Web Page Defacement Detection”, National Security Research Institute, Daejeon, Korea 2006.
- [2] Zone-h.org, [Online], Available: <http://www.zone-h.org/en/defacements> [Accessed 1-Dec-2014]
- [3] Bartoli Alberto, Davanzo Giorgio, and Medvet Eric “The Reaction Time to Website Defacements” IEEE INTERNET COMPUTING VOL. July/August (2009).
- [4] Paolo Passeri, [Online], “August 2014 Cyber Attacks Statistics”, Available: <http://hackmageddon.com/tag/august/> [Accessed 1 April 2015].
- [5] Peter Christen, “A Comparison of Personal Name Matching: Techniques and Practical Issues”, Department of Computer Science, The Australian National University, 2006.
- [6] Anna Kurniawati, Sulisty Puspitodjati, Sazali Rahman, “Implementasi Algoritma Jaro-Winkler Distance untuk Membandingkan Kesamaan Dokumen Berbahasa Indonesia”. Depok : Universitas Gunadarma, 2010.
- [7] William E. Yancey, “Evaluating String Comparator Performance for Record Linkage” Statistical Research Division U.S. Census Bureau. 2005.
- [8] TeSCA, “Webometrics, Peningkat Universitas Bergengsi”, [Online]. available : <http://tescaindonesia.org/content/news/webometrics-prestigious-university-rating-agency/read>. [Accessed : 17-Dec-2014]
- [9] Webometric, [Online], http://www.webometrics.info/en/About_Us, [Accessed: 13-Dec-2014]
- [10] Alexa, [Online], <http://www.alexa.com/about>, [Accessed : 14-Dec-2014]