

Analisis Kemiripan Struktur Label Website Pemerintah Daerah Menggunakan Algoritma *Greedy*

Arwendy Yunika Sari¹, Dana Sulisty Kusumo², Shinta Yulia Puspitasari³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹arwendyyunikasari@students.telkomuniversity.ac.id, ²danakusumo@telkomuniversity.ac.id,

³shinta1907@gmail.com

Abstrak

Standarisasi isi situs *website* pemerintah daerah telah dikeluarkan pada tahun 2003 namun ditahun-tahun sebelumnya *website* sudah mulai dibuat tetapi pada kenyataannya acuan standarisasi tersebut masih belum diterapkan. Pada penelitian sebelumnya didapatkan hasil bahwa masih terdapat *website* pemerintah daerah yang tidak sesuai dengan standarisasi konten yang sudah ditetapkan. Hal tersebutlah yang dapat menyebabkan adanya ketidaksesuaian informasi dengan standarisasi. Padahal fungsi adanya *website* yaitu untuk menyampaikan informasi yang valid kepada masyarakat terkait kebijakan pemerintah daerah yang berlaku. Acuan standarisasi yang digunakan yaitu berupa PANDUAN PEMBANGUNAN SITUS WEB PEMDA PESERTA USDRP. Manfaat yang dapat diperoleh pemerintah daerah dengan adanya penelitian ini yaitu dapat mengevaluasi kembali *website* mana yang menyediakan informasi yang sesuai dengan standarisasi berdasarkan hasil penelitian yang dilakukan. Perhitungan kemiripan dilihat dari *label/teks similarity* dan *structural similarity*. Dalam *label/teks similarity* akan dihitung kemiripan antar *string* dalam struktur label. Sedangkan dalam *structural similarity* struktur label akan direpresentasikan kedalam *graph* dan kemudian dihitung kemiripan *graph*nya. Dari hasil pengujian didapatkan bahwa algoritma *greedy* memiliki performansi waktu yang lebih cepat dibandingkan dengan algoritma *A** dalam hal memetakan *node* yang ada pada *graph*. Dalam melakukan iterasi, algoritma ini akan menghapus *node* yang sudah dipetakan sehingga waktu yang dibutuhkan untuk memetakan keseluruhan *node* menjadi lebih cepat dibandingkan dengan algoritma *A**. Hasil penelitian menunjukkan bahwa dari ketiga *website* (Tanjung Jabung Barat, Kuningan, dan Banjarnegera) yang memiliki kemiripan struktur label paling mirip terhadap Standarisasi adalah *website* Tanjung Jabung Barat dengan kemiripan sebesar 36,84%.

Kata kunci : *arsitektur informasi, string edit distance, string edit similarity, algoritma greedy, graph edit distance, graph edit similarity*

Abstract

The standardization of the contents of the website of the local government website was issued in 2003 but in previous years the website had begun to be made but in reality the standardization reference was still not implemented. In the previous research, it was found that there are still local government websites that are not in accordance with the predetermined content standardization. This is exactly what can cause information incompatibility with standardization. Whereas the function of the website is to convey valid information to the public related to the applicable local government policy. Standardization reference used is in the form of GUIDELINE DEVELOPMENT OF USDRP PARTICIPANTS WEB SITE. The benefits that can be obtained by local governments with this research is that they can re-evaluate which websites provide information in accordance with standardization based on the results of research conducted. Similarity calculation can be seen from the label / text similarity and structural similarity. In label / text similarity will be calculated similarity between strings in the label structure. Whereas in structural similarity the label structure will be represented into the graph and then the graph similarity is calculated.

From the test results it was found that the greedy algorithm has a faster time performance compared to the A * algorithm in terms of mapping existing nodes in the graph. In doing iteration, this algorithm will delete the mapped node so that the time needed to map the whole node becomes faster than the A * algorithm. The results showed that of the three websites (Tanjung Jabung Barat, Kuningan, and Banjarnegera) which had the most similar label structure similar to Standardization was the West Tanjung Jabung website with a similarity of 36.84%.

Keywords: *information architecture, string edit distance, string edit similarity, greedy algorithm, graph edit distance, graph edit similarity*

1. Pendahuluan

1. 2. Latar Belakang

Pembangunan *website* bagi Pemerintah Daerah merupakan implementasi dari Instruksi Presiden No. 3 Tahun 2003, tentang Kebijakan dan Strategi Nasional Pengembangan E-Government. Hal tersebut dilakukan untuk menunjang terwujudnya *good governance* di Indonesia. *Website* mulai digunakan sebagai salah satu media yang digunakan oleh pemerintah untuk menyalurkan informasi yang berisi berbagai macam konten terkait kebijakan pemerintah daerah yang berlaku kepada masyarakat.

Pemerintah sendiri telah mengeluarkan panduan isi pembangunan situs web pemerintah daerah pada tahun 2003 [3] padahal di tahun-tahun sebelumnya *website* sudah mulai dibuat. Pada kenyataannya, panduan yang telah dikeluarkan oleh DEPKOMINFO tidak dijadikan sebagai acuan dalam pembuatan dan pengembangan *website* [3]. Pada penelitian sebelumnya mengenai *website* pemerintah daerah kabupaten atau kota didapatkan hasil bahwa masih terdapat *website* yang tidak sesuai dengan usulan standarisasi konten yang dibuat [3]. Hal tersebut menyebabkan adanya ketidaksesuaian informasi terhadap standarisasi sehingga informasi yang disampaikan kepada masyarakat tidak valid. Standarisasi yang digunakan berupa buku Panduan Pembangunan Situs Web Pemda Peserta USDRP [6]. Berdasarkan latar belakang tersebut maka akan dilakukan penelitian dengan menghitung kemiripan struktur label antar *website* pemerintah daerah terhadap standarisasi dengan tujuan untuk mengetahui *website* pemerintah daerah mana yang memiliki kemiripan struktur label paling sesuai dengan standarisasi. Manfaat yang diperoleh pemerintah daerah dengan adanya penelitian ini yaitu pemerintah dapat melakukan evaluasi kembali terkait informasi yang ada dalam *website* setelah diperoleh nilai kemiripan strukturnya.

Penelitian dilakukan dengan menghitung *label/teks similarity (String Edit Distance dan String Edit Similarity)* untuk mendapatkan nilai kemiripan antar *string*. Setelah itu akan dilakukan perhitungan *structural similarity (Graph Edit Distance dan Graph Edit Simialrity)* yaitu dengan merepresentasikan struktur label ke dalam *graph* dan kemudian dihitung nilai kemiripan antar *graph* tersebut. Hasil akhir penelitian berupa nilai perbandingan kemiripan struktur label.

Dalam melakukan penelitian kemiripan *string* digunakan algoritma *greedy*, algoritma ini akan mencari atau memetakan pasangan *node* yang memiliki kemiripan *string* paling tinggi [11][17]. Selain itu, algoritma ini memiliki tingkat ketelitian lebih tinggi dibandingkan dengan algoritma *Exhaustive* [4][11] dan memiliki waktu eksekusi paling cepat dibandingkan dengan algoritma *A** dan *Exhaustive* [11] dan dapat digunakan dalam mencari solusi dari suatu permasalahan optimasi [12]. Oleh karena itu dalam penelitian ini digunakan algoritma *Greedy*.

1. 3. Rumusan Masalah

Berdasarkan latar belakang masalah diatas, rumusan masalah yang dapat diambil dalam penelitian ini yaitu:

- Bagaimana mengimplemantasikan algoritma *Greedy* dalam menghitung kemiripan struktur label pada *website* pemerintah daerah?
- Bagaimana hasil dan analisis data dari perhitungan kemiripan struktur label dengan menggunakan algoritma *Greedy*?

1. 4. Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini yaitu:

- Data yang digunakan adalah data label sebagai sistem navigasi (menu).
- Website* pemerintah daerah yang digunakan yaitu: Banjarnegara, Kuningan, dan Tanjung Jabung Barat.
- Hasil akhir penelitian berupa nilai kemiripan struktur label antar *website*.

1. 5. Tujuan

Tujuan dilakukannya penelitian ini yaitu:

- a) Melakukan implementasi algoritma *greedy* untuk menghitung kemiripan struktur label pada *website* pemerintah daerah.
- b) Melakukan analisis data dari hasil kemiripan label dan memberikan kesimpulan dari hasil analisis data yang diperoleh.

1. 6. Organisasi Tulisan

Penelitian ini tersusun dari beberapa bab, yaitu:

- a) Bab 1 : Pendahuluan
- b) Bab 2 : Studi Terkait
- c) Bab 3 : Gambaran Umum Sistem
- d) Bab 4 : Pengujian dan Analisis
- e) Bab 5 : Kesimpulan

2. Studi Terkait

2.2. Arsitektur Informasi

Arsitektur informasi merupakan suatu disiplin ilmu yang fokus untuk membuat informasi mudah ditemukan dan dipahami [2]. Komponen utama dalam arsitektur informasi terdiri dari: sistem *labeling*, sistem navigasi, sistem organisasi, dan sistem pencarian [2]. Dalam penelitian ini fokus dalam sistem *labeling*, sistem *labeling* yang dimaksud yaitu label sebagai sistem navigasi. Label sebagai sistem navigasi merepresentasikan pilihan menu yang ada dalam suatu *website*.

Dalam menghitung nilai kemiripan struktur label antar *website* dapat digunakan dengan menggunakan pendekatan *similarity of bussiness process model*. Untuk menghitung *similarity of bussiness process model*, hal pertama yang harus dilakukan yaitu dengan menghitung nilai *Label/Teks Similarity (String Edit Distance dan String Edit Similarity)* dan *Structure Similarity (Graph Edit Distance dan Graph Edit Similarity)* dimana struktur label harus direpresentasikan kedalam *graph* terlebih dahulu [14].

2.3. Label/Teks Similarity

Label/Teks Similarity digunakan untuk menghitung kemiripan antar *string*. Dalam perhitungan kemiripan *string* terdiri dari 2 proses, yaitu:

a) String Edit Distance

String Edit Distance atau *Levenshtein Distance* adalah minimal operasi yang dibutuhkan untuk mengubah *string* 1 menjadi *string* 2 [4]. Terdapat tiga macam operasi utama dalam metode ini, yaitu: operasi penyisipan (*insertion*), operasi penghapusan (*deletion*), dan operasi penggantian (*substitution*) [10]. Dalam perhitungan disediakan *tools Levenshtein Calculator*. Rumus yang digunakan yaitu [11]:

$$lev_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min \begin{cases} lev_{a,b}(i-1,j) + 1 \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1_{a_i \neq b_j} \end{cases} & \text{otherwise,} \end{cases}$$

b) String Edit Similarity

String Edit Similarity atau *ses* dilakukan untuk menghitung nilai kemiripan antara kedua *string* setelah sebelumnya dilakukan perhitungan *sed*. Nilai *similarity* yang diperoleh berada dalam skala 0-1 [14]. Rumus yang digunakan yaitu [8]:

$$Sim(n,m) = 1 - \frac{ed(l_1,l_2)}{\max(|l_1|,|l_2|)}$$

Keterangan :

$ed(l_1,l_2)$: nilai string edit distance

$\max(|l_1|,|l_2|)$: panjang string maksimum

2.4. Structural Similarity

Structural Similarity digunakan untuk menghitung kemiripan antar *graph*. Dalam perhitungan kemiripan *graph* terdiri dari 2 proses, yaitu:

a) Graph Edit Distance

Graph Edit Distance adalah minimal operasi yang dibutuhkan untuk mengubah suatu *graph* menjadi *graph* lain, operasi yang dilakukan yaitu: *insertion*, *deletions*, dan *substitutions node* [8][4]. Rumus *ged* yaitu [8]:

$$ged = |sn| + |se| + 2 \cdot \sum_{(n_1, n_2) \in M} (1 - Sim(n_1, n_2))$$

Untuk memperoleh nilai *ged* harus dilakukan perhitungan nilai *sn* dan *se*. *Sn* merupakan jumlah *node* yang di *insert* atau di *delete* dari *graph* [14], *Se* merupakan jumlah *edge* yang di *insert* atau di *delete* dari *graph* [14] dan *Sb* merupakan *node* pada *graph* 1 yang disubstitusikan ke *graph* 2 apabila *node* tersebut mirip. Berikut ini merupakan rumus yang digunakan:

$$|sn| = (N_1 \cup N_2) - \text{node substitusi on} \quad |se| = (E_1 \cup E_2) - \text{edge substitusi on}$$

$$Sb = \text{dom}(M) \cup \text{cod}(M)$$

Keterangan :

Se : jumlah insert/delete edge
 Sn : jumlah insert/delete node
 Sb : node substitution

b) Graph Edit Similarity

Graph Edit Similarity adalah nilai kemiripan antar *graph* namun untuk melakukan perhitungan harus dilakukan pemetaan untuk mendapatkan nilai kemiripan *node* yang maksimal. Nilai *similarity* yang diperoleh berada dalam skala 0-1 [14]. Rumus yang digunakan yaitu [8]:

$$simged(B1, B2) = 1 - avg(snv, sev, sbv)$$

Untuk menghitung *ges* harus dilakukan perhitungan *snv*, *sev*, dan *sbv* terlebih dahulu. *Snv* yaitu nilai rata-rata *node* yang di *insert/delete* dari *graph*, *Sev* yaitu nilai rata-rata *edge* yang di *insert/delete* dari *graph*, dan *Sbv* yaitu nilai rata-rata *node substitution*. Berikut ini merupakan rumus yang digunakan [8]:

$$snv = \frac{|sn|}{|N_1| + |N_2|}$$

$$sev = \frac{|se|}{|E_1| + |E_2|}$$

$$sbv = \frac{2 \cdot \sum_{(n,m) \in M} 1 - Sim(n, m)}{|N_1| + |N_2| - |sn|}$$

Keterangan :

Snv : nilai rata-rata insert/delete node
 Sn : jumlah insert/delete node
 N₁ : jumlah node graph 1
 N₂ : jumlah node graph 2
 Sev : nilai rata-rata insert/delete edge
 Se : jumlah insert/delete edge
 E₁ : jumlah edge graph 1
 E₂ : jumlah edge graph 2
 Sbv : nilai rata-rata node substitution

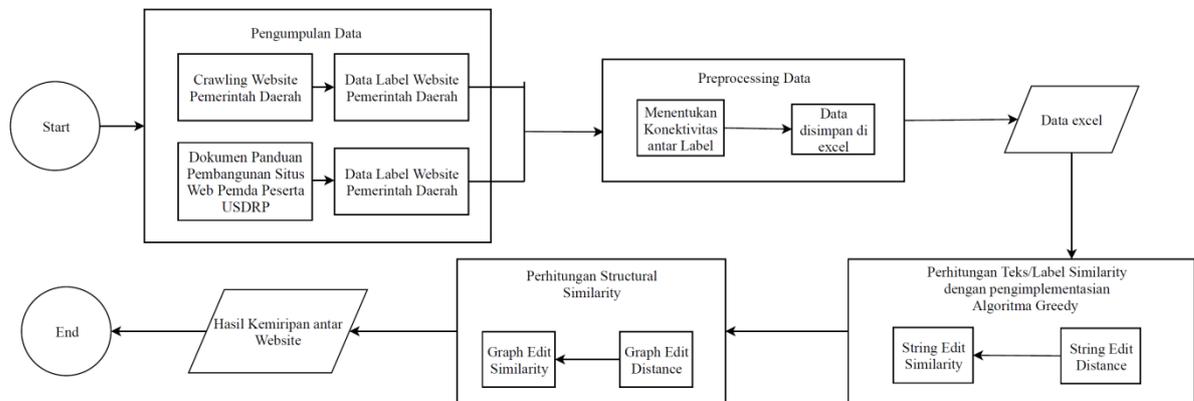
2. 5. Algoritma Greedy

Metode *Greedy* merupakan teknik yang biasa dilakukan untuk mencari solusi dari permasalahan optimasi, baik permasalahan maksimasi ataupun minimasi [12]. Algoritma ini akan menyusun secara *incremental mapping* antara dua pasang *graph* dan akan mulai menandai pasangan yang mungkin dari *node* berdasarkan dua *graph* sebagai *open pairs* [4]. Dari kemungkinan pasangan *node* yang mungkin, akan dicari nilai *string edit similarity* tertinggi [14]. Apabila pasangan *node* tersebut sudah terpilih menjadi pasangan dengan nilai *string edit similarity* tertinggi maka *node* tersebut akan masuk ke dalam pemetaan, dan *node* tersebut akan dihapus sebagai *open pairs*. Hal tersebut dilakukan agar *node* yang sudah masuk ke dalam pemetaan tidak dipilih lagi pada iterasi berikutnya. Iterasi akan tetap berlangsung sampai tidak ada *node* yang bisa dipasangkan lagi.

3. Sistem yang Dibangun

3.2. Gambaran Umum Sistem

Gambar 1 berikut ini merupakan gambaran umum sistem yang dibangun dalam penelitian ini:



Gambar 1. Gambaran Umum Sistem

Tahapan yang dilakukan berdasarkan gambar 1 diatas adalah:

- 1) Pengumpulan data dilakukan dengan tujuan untuk memperoleh data label *website* dan standarisasi. Data label *website* pemerintah daerah diperoleh dengan melakukan *crawling* data menggunakan *Webscompare* [13] sedangkan untuk label standarisasi diperoleh dengan melakukan *input* data manual ke excel.
- 2) *Preprocessing* data dilakukan dengan tujuan untuk memperoleh konektivitas antar label yang telah di *crawling* sebelumnya. Konektivitas antar label menunjukkan level menu yang ada pada *website*. Tahap *preprocessing* ini dilakukan secara manual dan kemudian disimpan dalam file excel.
- 3) Melakukan perhitungan *String Edit Distance* dan *String Edit Similarity* antar label dengan tujuan untuk memperoleh nilai kemiripan *string*.
- 4) Melakukan pemetaan dengan algoritma *Greedy*, tujuannya yaitu untuk memperoleh nilai kemiripan *string* yang paling maksimum berdasarkan pemetaan yang telah dilakukan.
- 5) Melakukan perhitungan *Graph Edit Distance* dan *Graph Edit Similarity* dengan tujuan untuk memperoleh nilai kemiripan antar *graph*.
- 6) Melakukan analisis data berdasarkan nilai kemiripan *graph* yang diperoleh pada tahap sebelumnya.

3.3. Pengumpulan dan Preprocessing Data

Tahap pengumpulan data *website* dilakukan dengan melakukan *crawling* menggunakan *Webscompare* sedangkan untuk data standarisasi diperoleh dengan melakukan *input* data manual. Setelah data label diperoleh maka tahap yang dilakukan setelahnya yaitu menentukan konektivitas antar label secara manual. Data tersebut kemudian disimpan ke dalam file excel. Untuk representasi *graph* konektivitas antar menu Standarisasi dapat dilihat pada Lampiran Gambar 4 dan konektivitas antar menu Banjarnegara dapat dilihat pada Lampiran Gambar 5. Berikut ini merupakan salah satu contoh tahap pengumpulan data dan *preprocessing* data yang disimpan ke dalam file excel:

Label	Konektivitas 2	Konektivitas 1
Selayang Pandang		Selayang Pandang
Sejarah		Selayang Pandang
Motto Daerah		Selayang Pandang
Lambang dan Arti Lambang		Selayang Pandang
Lokasi		Selayang Pandang
Visi dan Misi		Selayang Pandang
Pemerintahan Daerah		Pemerintah Daerah
Struktur Organisasi		Pemerintah Daerah
Nama		Pemerintah Daerah
Alamat		Pemerintah Daerah
Telpon		Pemerintah Daerah
E-mail		Pemerintah Daerah
Geografi		Geografi
Topografi		Geografi
Demografi		Geografi
Cuaca dan Iklim		Geografi
Sosial dan Ekonomi		Geografi
Budaya		Geografi
Peta Wilayah dan Sumberdaya		Peta Wilayah dan Sumberdaya
Peraturan/Kebijakan Daerah		Peraturan/Kebijakan Daerah
Berita		Berita
Berita Pemda		Berita
Forum Diskusi		Forum Diskusi
Saran/Komentar Pengujung pada Buku Tamu		Saran/Komentar Pengujung pada Buku Tamu
Program/Kegiatan Pembangunan		Program/Kegiatan Pembangunan
Dokumen RPJMD		Dokumen RPJMD
Ringkasan APBD		Ringkasan APBD
Laporan Audit Tahunan		Laporan Audit Tahunan
Daftar Paket Kontrak pada tahun berjalan		Daftar Paket Kontrak pada tahun berjalan
Daftar Paket Terkontrak		Daftar Paket Terkontrak

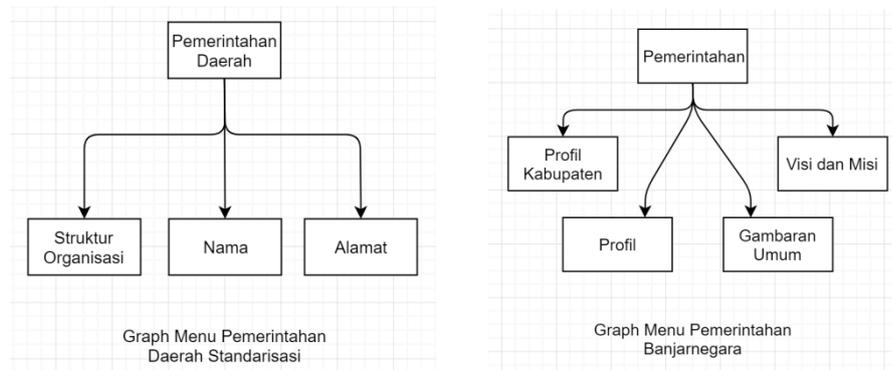
Gambar 2. Konektivitas label Standarisasi

Label	Konektivitas 2	Konektivitas 1
Home	Home	Home
Sambutan	Sambutan	Sambutan
Pemerintahan	Pemerintahan	Pemerintahan
Profil Kabupaten	Pemerintahan	Pemerintahan
Profil	Pemerintahan	Pemerintahan
Visi dan Misi	Pemerintahan	Pemerintahan
Gambaran Umum	Pemerintahan	Pemerintahan
Letak Geografis	Pemerintahan	Pemerintahan
Adat Istiadat	Pemerintahan	Pemerintahan
Sarana dan Prasarana	Pemerintahan	Pemerintahan
Pemerintahan	Pemerintahan	Pemerintahan
Arti Lambang Daerah	Pemerintahan	Pemerintahan
Semboyan	Pemerintahan	Pemerintahan
Kelembagaan	Pemerintahan	Pemerintahan
Sekretariat Daerah	Pemerintahan	Pemerintahan
Bagian Pemerintahan dan Otda	Sekretariat Daerah	Pemerintahan
Bagian Hukum	Sekretariat Daerah	Pemerintahan
Bagian Kesra	Sekretariat Daerah	Pemerintahan
Bagian Perekonomian	Sekretariat Daerah	Pemerintahan
Bagian Pembinaan BUMD	Sekretariat Daerah	Pemerintahan
Bagian Pembangunan	Sekretariat Daerah	Pemerintahan
Bagian Organisasi	Sekretariat Daerah	Pemerintahan
Bagian Umum	Sekretariat Daerah	Pemerintahan
Bagian Keuangan	Sekretariat Daerah	Pemerintahan
Sekretariat DPRD	Pemerintahan	Pemerintahan
Inspektorat	Pemerintahan	Pemerintahan
Dinas	Pemerintahan	Pemerintahan
Dindikpora	Dinas	Pemerintahan
Dinkes	Dinas	Pemerintahan
Dinsos	Dinas	Pemerintahan

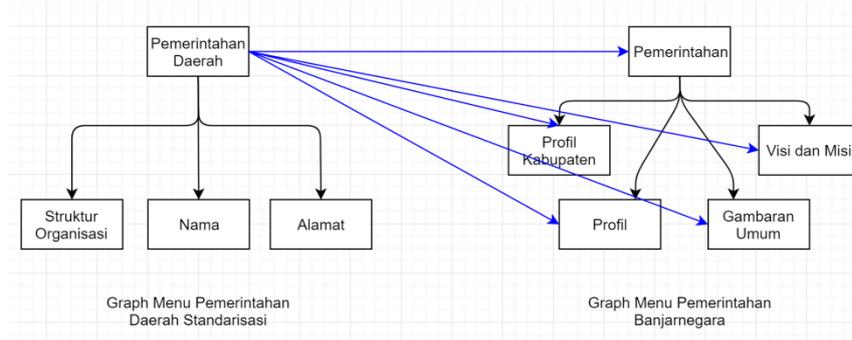
Gambar 3. Konektivitas label Banjarnegara

3.4. Perhitungan String Edit Distance dan String Edit Similarity

Berikut ini merupakan salah satu contoh perhitungan *String Edit Distance* dan *String Edit Similarity* yang sebelumnya data labelnya diubah ke dalam *graph*:



Sebelum dilakukan perhitungan *String Edit Distance* dan *String Edit Similarity*, terlebih dahulu harus dilakukan pemetaan untuk setiap *node* yang ada. Berikut ini merupakan contoh pemetaan yang dilakukan berdasarkan gambar diatas:



Tanda panah biru menunjukkan kemungkinan pasangan *node* yang akan dipetakan. Kemungkinan pasangan *node* terdiri dari : {(Pemerintahan Daerah, Pemerintahan), (Pemerintahan Daerah, Profil Kabupaten), (Pemerintahan Daerah, Profil), (Pemerintahan Daerah, Visi dan Misi), (Pemerintahan Daerah, Gambaran Umum)}. Dari kemungkinan pasangan *node* tersebut kemudian akan dihitung nilai *String Edit Distance* dan *String Edit Similarity*, berikut ini merupakan perhitungan yang dilakukan:

- (Pemerintahan Daerah, Pemerintahan)

$$sed = 7, ses = 1 - (7/19) = \mathbf{0,6315}$$

- (Pemerintahan Daerah, Profil Kabupaten)

$$sed = 7, ses = 1 - (15/19) = 0,2105$$

- (Pemerintahan Daerah, Profil)

$$sed = 7, ses = 1 - (17/19) = 0,1052$$

- (Pemerintahan Daerah, Visi dan Misi)

$$sed = 7, ses = 1 - (15/19) = 0,2105$$

- (Pemerintahan Daerah, Gambaran Umum)

$$sed = 7, ses = 1 - (14/19) = 0,2631$$

Berdasarkan nilai *ses*/kemiripan *string* yang diperoleh maka pasangan yang masuk ke dalam pemetaan adalah {(Pemerintahan Daerah, Pemerintahan)}. Pasangan tersebut dipetakan karena memiliki nilai kemiripan *string* tertinggi dibandingkan dengan pasangan lainnya.

3. 5. Pemetaan dengan Algoritma Greedy

Dalam penelitian ini dilakukan pengujian terkait performansi waktu algoritma *greedy*. Pengujian dilakukan dengan membandingkan algoritma *greedy* dan algoritma A^* . Cara kerja algoritma *greedy* yaitu dengan memetakan pasangan *node* yang memiliki nilai kemiripan *string* paling maksimum untuk setiap iterasinya, yang kemudian di iterasi berikutnya pasangan *node* tersebut akan dihapus sehingga tidak dipetakan lagi. Sedangkan algoritma A^* memiliki cara kerja dengan memetakan keseluruhan *node* yang memiliki nilai kemiripan *string* lebih besar atau sama dengan *cut off value* yang telah ditentukan sebelumnya [18][19].

Pengukuran waktu menggunakan algoritma A^* diperoleh dengan melakukan pengujian antara data label Standarisasi dengan Banjarnegera dan kemudian dihitung lama waktu yang dibutuhkan dalam memetakan *node* yang ada. Setelah itu, hasil pengukuran waktu yang diperoleh algoritma A^* akan dibandingkan dengan hasil pengukuran waktu yang diperoleh algoritma *greedy*. Dari hasil perbandingan maka akan diperoleh hasil algoritma mana yang lebih unggul terkait performansi waktunya. Algoritma A^* diperoleh berdasarkan penelitian yang telah dilakukan oleh peneliti sebelumnya [20].

3. 6. Perhitungan Graph Edit Distance dan Graph Edit Similarity

Untuk melakukan perhitungan *Graph Edit Distance* dan *Graph Edit Similarity*, data yang diperlukan yaitu nilai *ses* tertinggi yang diperoleh berdasarkan hasil pemetaan yang telah dilakukan sebelumnya. Pengimplementasian algoritma *greedy* dilakukan untuk melakukan pemetaan terhadap pasangan *node* yang ada. Apabila *node* pada *graph* 1 sudah dipetakan dengan *node* pada *graph* 2 maka pada iterasi berikutnya *node* tersebut akan dihapus dari *graph*, sehingga pada iterasi berikutnya *node* tersebut tidak dapat dipetakan lagi. Proses iterasi akan terus berlangsung sampai semua *node* di *graph* 1 dipetakan dengan *graph* 2. Algoritma ini dipilih dalam penelitian karena algoritma ini pernah digunakan dalam penelitian *graph* yang sama [11].

Perhitungan *Graph Edit Distance* dan *Graph Edit Similarity* dilakukan berdasarkan contoh yang sebelumnya. Hasil pemetaan yang diperoleh yaitu: {(Pemerintahan Daerah, Pemerintahan), (Struktur Organisasi, Visi dan Misi), (Nama, Gambaran Umum), (Alamat, Profil Kabupaten)}. Berikut ini merupakan rumus *ged* yang digunakan:

$$ged = |sn| + |se| + 2 \cdot \sum_{(n,m) \in M} 1 - (Sim(n, m))$$

Dimana:

$$|sn| = (N_1 \cup N_2) - \text{Node Substitutions (sb)} \quad |se| = (E_1 \cup E_2) - \text{edge substitutions}$$

Berdasarkan hasil pemetaan, maka diperoleh nilai perhitungan sebagai berikut ini:

$$\begin{aligned} sn &= 5, \quad se = 1 \\ ged &= 5 + 1 + 2 \cdot ((1 - 0,6315) + (1 - 0,2631) + (1 - 0,2307) + (1 - 0,25)) \\ &= 6 + 2 \cdot (2,6247) \\ &= 6 + 5,2494 \\ &= 11,7694 \end{aligned}$$

Setelah memperoleh nilai *ged*, maka tahap yang dilakukan yaitu menghitung nilai *ges*. Rumus yang digunakan yaitu:

$$ges = 1 - \frac{snv + sev + sbv}{3}$$

Dimana:

$$snv = \frac{|sn|}{|N_1| + |N_2|} \quad sbv = \frac{2 \cdot \sum_{(n,m) \in M} 1 - (Sim(n, m))}{|N_1| + |N_2| - |sn|} \quad sev = \frac{|se|}{|E_1| + |E_2|}$$

Berdasarkan nilai *se* dan *sn* yang diperoleh sebelumnya, maka perhitungan dengan rumus diatas dapat dilakukan seperti berikut ini:

$$\begin{aligned}
 snv &= 5 / (4+5) \\
 &= 5/9 \\
 &= 0,5555 \\
 sbv &= (2 \cdot ((1 - 0,6315) + (1 - 0,2631) + (1 - 0,2307) + (1 - 0,25))) / (4+5-5) \\
 &= (2 \cdot (2,6247)) / 4 \\
 &= (11,7694) / 4 \\
 &= 2,9423 \\
 sev &= 1 / (3+4) \\
 &= 1/7 \\
 &= 0,1428 \\
 ges &= 1 - ((0,5555 + 0,1428 + 2,9423) / 7) \\
 &= 1 - (3,6406 / 7) \\
 &= 1 - 0,52 \\
 &= 0,48
 \end{aligned}$$

Berdasarkan nilai *ges* yang diperoleh, dapat disimpulkan bahwa nilai kemiripan *website* pemerintah daerah Bangka terhadap Standarisasi adalah sebesar 0,48 atau 48%.

4. Evaluasi

4.2. Skenario Pengujian

Skenario pengujian dalam penelitian ini terdiri dari:

1) Melakukan perhitungan *String Edit Distance* dan *String Edit Similarity* antar label Standarisasi dengan label *website* pemerintah daerah, dengan tujuan untuk mengetahui nilai kemiripan antar *string*. Pengujian dilakukan dengan membandingkan nilai *output* program dengan nilai *Levensthein Calculator*. Apabila dari hasil pengujian nilai yang diperoleh adalah sama maka dapat dikatakan bahwa nilainya valid.

2) Melakukan pengujian dengan mengimplementasikan algoritma *greedy* dalam memetakan *node* dari *graph* 1 (Standarisasi) ke dalam *graph* 2 (Banjarnegara). Tujuan dilakukannya pengujian ini yaitu untuk menemukan perbandingan waktu yang dibutuhkan antara algoritma *greedy* dengan algoritma A^* dalam menghitung kemiripan struktur label yaitu dalam melakukan pemetaan *node*.

3) Melakukan perhitungan *Graph Edit Distance* dan *Graph Edit Similarity* antar struktur label Standarisasi dengan *website* pemerintah daerah, dengan tujuan untuk mengetahui nilai kemiripan struktur *website* sudah sama. Sama disini maksudnya yaitu nilai *output* program sama dengan nilai perhitungan secara *tracing* manual.

4.3. Pengujian String Edit Distance dan String Edit Similarity terhadap Levensthein Distance

Dalam pengujian ini, sampel data yang digunakan yaitu data label standarisasi dan data label banjarnegara. Tujuan dilakukannya pengujian ini yaitu untuk menunjukkan bahwa *output* program sama dengan *Levensthein Calculator* sehingga *output* yang diperoleh adalah valid. Setelah dilakukan pengujian antara standarisasi dengan banjarnegara diperoleh hasil seperti Tabel 1 dibawah ini:

Tabel 1. Perbandingan String Edit Distance dan String Edit Similarity antara program dengan Levensthein Calculator (Standarisasi terhadap Banjarnegara)

Data Label	Program	Levensthein Calculator	Keterangan
------------	---------	------------------------	------------

Standarisasi	Banjarnegara	sed	ses	sed	ses	
Selayang Pandang	Pelayanan	8	0.5	8	0,5	Valid
Sejarah	Semboyan	5	0.375	5	0,375	Valid
Motto Daerah	Sekretaris Daerah	9	0.47058 8235	9	0,4705	Valid
Lambang dan Arti Lambang	Sarana dan Prasarana	12	0.5	12	0,5	Valid
Lokasi	Dinas	4	0.33333 3333	4	0,3333	Valid
Visi dan Misi	Visi dan Misi	0	1	0	1	Valid
Pemerintahan Daerah	Pemerintahan	7	0.63157 8947	7	0,6315	Valid
Struktur Organisasi	Bagian Organisasi	8	0.57894 7368	8	0,5789	Valid
Nama	Sambutan	5	0.375	5	0,375	Valid
Alamat	Anggaran	5	0.375	5	0,375	Valid

Berdasarkan Tabel 1 diatas, dapat disimpulkan bahwa semua *output* data bernilai valid. Hal tersebut telah dibuktikan dari hasil perbandingan *output* program yang diperoleh adalah sama dengan *Levensthein Calculator*. Untuk hasil perbandingan perhitungan kemiripan *string* antara Standarisasi dengan Banjarnegara secara keseluruhan dapat dilihat pada Lampiran Tabel 5.

4. 4. Pengujian Algoritma Greedy

Dalam pengujian ini, sampel data yang digunakan yaitu data label Standarisasi dan Banjarnegara. Pengujian dilakukan dengan membandingkan algoritma *greedy* dengan algoritma *A**. Tujuan dilakukannya pengujian ini yaitu untuk menemukan perbandingan waktu yang dibutuhkan oleh algoritma *greedy* dan algoritma *A** dalam menghitung kemiripan struktur label dengan melakukan pemetaan *node*. Pengujian waktu algoritma *greedy* dan algoritma *A** dilakukan dengan membandingkan data label Standarisasi dengan Banjarnegara yang kemudian dihitung berapa lama waktu yang dibutuhkan untuk memetakan keseluruhan *node* yang ada. Berikut ini merupakan Tabel 2 hasil pengujian yang telah dilakukan:

Tabel 2. Hasil pengujian performansi waktu Algoritma Greedy dengan *A**

Algoritma Greedy				Algoritma <i>A*</i>			
Standarisasi	Banjarnegara	ses	Waktu	Standarisasi	Banjarnegara	ses	Waktu
Selayang Pandang	Pelayanan	0.5	01:75	Selayang Pandang	Pelayanan	0.5000	02:92
Sejarah	Semboyan	0.375		Sejarah	Pengaduan	0.3333	
Motto Daerah	Sekretaris Daerah	0.4705 88235		Motto Daerah	Sekretaris Daerah	0.4706	
Lambang dan Arti Lambang	Sarana dan Prasarana	0.5		Lambang dan Arti Lambang	Sarana dan Prasarana	0.5000	

Lokasi	Dinas	0.3333 33333	Lokasi	Dinas	0.3333
Visi dan Misi	Visi dan Misi	1	Visi dan Misi	Visi dan Misi	1
Pemerintahan Daerah	Pemerintahan	0.6315 78947	Pemerintahan Daerah	Pemerintahan	0.6316
Struktur Organisasi	Bagian Organisasi	0.5789 47368	Struktur Organisasi	Bagian Organisasi	0.5789
Nama	Sambutan	0.375	Nama	Sambutan	0.3750
Alamat	Anggaran	0.375	Alamat	Adat Istiadat	0.3077

Berdasarkan Tabel 2 diatas, dapat dilihat bahwa performansi (waktu yang dibutuhkan untuk melakukan pemetaan) algoritma *greedy* lebih cepat dibandingkan dengan algoritma A^* . Algoritma *greedy* membutuhkan waktu selama 1 menit 75 detik untuk melakukan pemetaan antara standarisasi dan banjarnegara dengan jumlah *node* sebanyak 30 *node* dibandingkan dengan 92 *node*, sedangkan untuk algoritma A^* membutuhkan waktu selama 2 menit 92 detik untuk memetakan 30 *node* pada standarisasi terhadap 92 *node* banjarnegara. Hal tersebut dapat dibuktikan pada Lampiran Tabel 6. Pada algoritma *greedy* untuk setiap iterasi *node* yang sudah dipetakan dengan kemiripan *string* paling tinggi akan dihapus sehingga pada iterasi berikutnya *node* tersebut tidak akan dihitung lagi kemiripan stringnya sedangkan pada algoritma A^* akan memetakan *node* secara keseluruhan terlebih dahulu dan kemudian akan dipetakan kembali berdasarkan *cut off value* yang digunakan, hal tersebutlah yang menyebabkan algoritma *greedy* memiliki performansi waktu lebih cepat dibanding dengan algoritma A^* . Berdasarkan pengujian tersebut dapat disimpulkan bahwa algoritma *greedy* memiliki performansi waktu lebih cepat dibandingkan dengan algoritma A^* [11].

Tabel 3. Hasil perbandingan kemiripan graph antara Standarisasi dengan Banjarnegara menggunakan Algoritma Greedy dan Algoritma A^*

Website 1	Website 2	Algoritma Greedy	Algoritma A^*
		ges	ges
Standarisasi	Banjarnegara	0,2088	0,2313

Berdasarkan Tabel 3 diatas dapat dilihat perbandingan nilai kemiripan *graph* (Standarisasi terhadap Banjarnegara) yang diperoleh dengan algoritma *greedy* dan A^* . Nilai kemiripan *graph* yang diperoleh menggunakan algoritma *greedy* adalah 0,2088 sedangkan nilai kemiripan *graph* yang diperoleh menggunakan algoritma A^* adalah sebesar 0,2313. Berdasarkan hasil yang diperoleh dapat dilihat bahwa nilai kemiripan *graph* yang dihasilkan algoritma A^* lebih tinggi daripada algoritma *greedy*.

4. 5. Pengujian Graph Edit Distance dan Graph Edit Similarity

Pada pengujian ini, sampel data yang digunakan yaitu struktur label Standarisasi dan Banjarnegara. Pengujian kevalidan nilai kemiripan *graph* akan dilakukan dengan membandingkan struktur label Standarisasi terhadap Standarisasi. Apabila dilakukan pengujian dengan data yang sama maka nilai kemiripan *graph* yang diperoleh adalah sebesar 1 atau 100%. setelah nilai kemiripan yang dihasilkan sudah valid, maka pengujian antara standarisasi dengan ketiga *website* dapat dikatakan valid juga nilainya. Tabel 4 berikut ini merupakan tabel hasil pengujian yang telah dilakukan:

Tabel 4. Hasil perbandingan kemiripan graph antara standarisasi terhadap ketiga website pemerintah daerah

Struktur Label		Program
Website 1	Website 2	ges
Standarisasi	Standarisasi	1
Standarisasi	Banjarnegara	0,2088
Standarisasi	Kuningan	0,0743
Standarisasi	Tanjung Jabung Barat	0,3684

Berdasarkan Tabel 4, dapat dilihat bahwa kemiripan *graph* yang diperoleh dari perbandingan Standarisasi dengan Standarisasi adalah 1, Standarisasi dengan Banjarnegara sebesar 0,2088, Standarisasi dengan Kuningan sebesar 0,0743, dan Standarisasi dengan Tanjung Jabung Barat sebesar 0,3684.

Untuk melakukan validasi terhadap hasil yang diperoleh pada Tabel 4, peneliti melakukan pengujian terhadap *node* yang tidak sesuai dengan Standarisasi. Untuk hasil pengujian dapat dilihat pada Lampiran Tabel 7. *Website* Tanjung Jabung Barat memiliki kemiripan tertinggi terhadap Standarisasi, hal tersebut ditunjukkan dengan banyaknya jumlah *node* yang tidak sesuai Standarisasi sebanyak 25 *node* dari total 31 *node*. Jumlah tersebut paling sedikit dibandingkan dengan kedua *website* lainnya. Untuk *website* Kuningan memiliki kemiripan *graph* terendah terhadap Standarisasi, hal tersebut dibuktikan dengan banyaknya jumlah *node* yang tidak sesuai dengan Standarisasi sebanyak 109 *node* dari total 125 *node*. Sedangkan *website* Banjarnegara memiliki kemiripan dibawah *website* Tanjung Jabung Barat namun diatas *website* Kuningan yaitu dengan jumlah *node* yang tidak sesuai Standarisasi sebanyak 82 *node* dari total 92 *node*.

Berdasarkan Tabel 4 dan Lampiran Tabel 7 dapat disimpulkan bahwa kemiripan *graph* dipengaruhi oleh banyaknya jumlah *node* yang tidak sesuai dengan standarisasi. Semakin banyak jumlah *node* yang tidak sesuai dengan standarisasi maka semakin rendah juga nilai kemiripan *graph* yang dihasilkan. Hasil pengujian menunjukkan bahwa hasil yang diperoleh adalah sama sehingga dapat dikatakan bahwa hasil penelitian valid.

5. Kesimpulan

Berdasarkan pengujian yang dilakukan dapat disimpulkan bahwa algoritma *greedy* memiliki performansi waktu yang lebih cepat dibandingkan dengan algoritma A^* . Performansi waktu yang dimaksud disini adalah waktu yang dibutuhkan untuk memetakan keseluruhan *node* pada *graph*.

Dari hasil pengujian diperoleh *website* Banjarnegara dengan nilai 0.2088, *website* Kuningan dengan nilai 0.0743, dan *website* Tanjung Jabung Barat dengan nilai 0.3684. Berdasarkan hasil pengujian tersebut maka dapat disimpulkan bahwa dari ketiga *website* (Banjarnegara, Kuningan, dan Tanjung Jabung Barat) yang memiliki kemiripan struktur label paling mirip dengan Standarisasi adalah *website* Tanjung Jabung Barat dengan nilai kemiripan struktur label sebesar 0.3684 atau apabila diubah menjadi persentase yaitu sebesar 36,84%. Kemiripan struktur label sendiri dipengaruhi oleh banyaknya jumlah *node* yang tidak sesuai dengan Standarisasi, semakin banyak *node* yang tidak sesuai dengan Standarisasi maka kemiripan struktur labelnya akan semakin rendah. Dalam hal performansi waktu algoritma *greedy* lebih unggul dibandingkan dengan algoritma A^* , hal tersebut dikarenakan dalam setiap iterasinya algoritma *greedy* akan memetakan *node* dengan nilai kemiripan *graph* paling tinggi dan kemudian dihapus sehingga di iterasi berikutnya tidak dihitung lagi. Sedangkan pada algoritma A^* akan menghitung kemiripan *graph* untuk keseluruhan *node* yang ada dahulu baru kemudian akan dipetakan berdasarkan cut off valuenya.

Daftar Pustaka

- [1] A. Supriyanto, "Arsitektur dan Teknik Desain Informasi pada Web," vol. XII, no.1, pp. 1–9, 2007.
- [2] P. Morville, L. Rosenfeld J.Arango. 2006. *Information architecture for the world wide web fourth edition*. United State : O'Reilly Media, Inc.
- [3] N. Herlan Puspa, "Analisis Kajian Standarisasi Isi Situs Web Pemerintah Daerah Kabupaten/Kota," 2018. [Online]. Available: <http://jurnal.upnyk.ac.id/index.php/semnasif/article/view/755>.
- [4] Dijkman, R., Dumas, M., & García-Bañuelos, L. (2009). Graph Matching Algorithms for Business Process Model Similarity Search. *Business Process Management*, 48–63.
- [5] "Website Kabupaten Terbaik," 2019. [Online]. Available: <http://egovbench.addi.is.its.ac.id/>.
- [6] Studylibid.com, "PANDUAN PEMBANGUNAN SITUS WEB PEMDA PESERTA USDRP," 2018. [Online]. Available: <https://studylibid.com/doc/348510/panduan-pembangunan-situs-web-pemda-peserta-usdrp>.
- [7] H. L. Z. Robertus Theodore, "Analisis Model Ekstraksi Data Pada Halaman Web.pdf," *Pros. Konf. Nas. Inform.*, p. 175, 2013.
- [8] Dijkman, R., Dumas, M., & Garcia-Banuelos, L. (2012). Business Process Graphs: Similarity Search and Matching. *Graph Data Management: Techniques and Applications* (p. 421-432). Hershey, PA: IGI Global.
- [9] R. Haldar and D. Mukhopadhyay, "Levenshtein Distance Technique in Dictionary Lookup Methods: An Improved Approach," *arXiv:1101.1232*, no. Ld, pp. 286–293, 2011.
- [10] Ariyani N.H, Sutardi & Rmadhan R, "Aplikasi Pendeteksi Kemiripan Isi Teks Dokumen Menggunakan Metode Levenshtein Distance," 2018. [Online]. Available: <http://ojs.uho.ac.id/index.php/semantik/article/view/1030>.
- [11] M. D. L. G.-B. Remco Dijkman, "Graph Matching Algorithm for Business Process Model Similarity Search," 2009.
- [12] Bednorz, W. (2008). *Greedy Algorithms*. Rijek, Croatia: InTech
- [13] "Web Menu Crawler", 2019. [Online]. Available: <http://webscompare.com/>.
- [14] R. M. Dijkman, M. Dumas, B. F. van Dongen, R. Kaarik, and J. Mendling. *Similarity of business process models: Metrics and evaluation*. Working Paper 269, BETA Research School, Eindhoven, The Netherlands, 2009.
- [15] Faizal, Muhammad, "Analisis dan Implementasi Graph Edit Distance dengan Algoritma Greedy untuk Similarity Search dalam Business Process Graphs," 2016. [Online]. Available: <https://openlibrary.telkomuniversity.ac.id/pustaka/files/115242/jurnal/analisis-dan-implementasi-graph-edit-distance-dengan-algoritma-greedy-untuk-similarity-search-dalam-business-process-graphs.pdf>
- [16] Dwiyantri, Fadhillah, "Analisis Bussiness Process Model Similarity Checking Menggunakan Teknik Greedy Graph Matching," 2015. [Online]. Available: <https://id.123dok.com/document/zx9o77nz-analisis-business-process-model-similarity-checking-menggunakan-teknik-greedy-graph-matching.html>
- [17] R. Dijkman, M. Dumas, L. Garcia-Banuelos dan R. Kaarik, "Aligning Business Process Models," dalam *Enterprise Distributed Object Computing Conference, 2009. EDOC '09. IEEE International*, Auckland, 2009.
- [18] Ventrinur, Anisyadita, "Analisis Kemiripan Struktural Model Proses Bisnis Dengan Implementasi A* Graph Matching," 2015.
- [19] S. D. Sembiring, "Analisis Kemiripan Model Proses Bisnis menggunakan Algoritma Heuristik," 2016. [Online]. Available: <https://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/1222/1167>
- [20] Anis, Amelia, "Analisis Kemiripan Struktur Label Website Pengadilan Tinggi di Indonesia dengan Implementasi A* Graph Matching", 2019.