

ANALISIS IMPLEMENTASI *SYNTATIC LABEL MATCHING SIMILARITY* UNTUK MEMERIKSA KEMIRIPAN LABEL PADA MODEL PROSES BISNIS

ANALYSIS IMPLEMENTATION *SYNTATIC LABEL MATCHING SIMILARITY* FOR CHECKING LABEL SIMILARITY OF BUSINESS PROCESS MODEL

Angdini Dea Nurillailiy¹, Dana Sulistyio Kusumo, ST., MT., P.hD², Angelina Prima Kurniati, ST., MT.³

^{1,2,3}Prodi S1 Teknik Informatika, Fakultas Teknik, Universitas Telkom

deanurillaiy@gmail.com, ² dskusumo.itt@gmail.com, ³angelina@telkomuniversity.ac.id

Abstrak

Similarity checking ini dilakukan untuk mencari tingkat kemiripan dari dua proses bisnis yang dibandingkan. Dalam melakukan pencarian kemiripan proses bisnis memiliki tiga metode yang digunakan, berdasarkan label, perilaku, dan struktural. Pada penelitian ini, dilakukan pengecekan kemiripan proses bisnis berdasarkan label.

Pada *label similarity* ada beberapa teknik yaitu *syntatic* dan *semantic*. Penelitian ini berfokus pada *syntatic similarity*. Pengecekan ini melakukan perhitungan berdasarkan sintaks. Untuk dapat melakukan perhitungan tersebut, maka dapat dilakukan *mapping* untuk pencarain kata yang mirip pada kedua proses bisnis tersebut. Selanjutnya, untuk melakukan pencarian tingkat kemiripan proses bisnis dapat dilakukan perhitungan *label matching similarity*. Pemodelan yang dilakukan untuk menggambarkan proses bisnis adalah *Business Process Modelling and Notation (BPMN)*.

Hasil perbandingan yang dilakukan dengan metode *syntatic label matching similarity* menghasilkan nilai perbandingan antara Reklame dan Usaha Bidang Penanaman Modal mendapatkan nilai 98,66%, perbandingan antara Reklame dan Surat Tanda Daftar Perusahaan 80% dan perbandingan Surat Tanda Daftar Perusahaan dan Usaha Bidang Penanaman Modal adalah 72,72%.

Kata Kunci : Proses Bisnis, *Similarity Checking*, *Label Similarity*, *Syntatic Similarity*, *Label Matching Similarity*, BPMN.

Abstract

Similarity checking is used to find of similarity between two business processes compared. In similarity checking has three methods are used, based on label, behavioral, and structural. In this research, similarity checking use based on label.

Label similarity has three techniques is syntatic, semantic and contextual. This research focus on syntatic similarity. Checking similarity use calculations based syntac. To calculate syntac, it can be mapped to find similar word between business process. After these calculations, to find similarity degree of business process we can do calculations label matching similarity. Modelling was done to describe the business process is Business Process Modelling and Notation (BPMN).

Result the compare two business process with syntatic label matching similarity between 'Reklame' and 'Usaha Bidang Penanaman Modal' get score similarity 98,66%, 'Reklame' and 'Surat Tanda Daftar Perusahaan' get score similarity 80% and 'Surat Tanda Daftar Perusahaan' and 'Usaha Bidang Penanaman Modal' get score similarity 72,72%.

Keywords : Business Process, Similarity checking, Label similarity, Syntatic similarity, Label matching similarity, BPMN

1. Pendahuluan

Proses bisnis sudah menjadi hal yang biasa untuk sebuah organisasi dikarenakan proses bisnis ini dibuat untuk mencapai visi dan misi dan organisasi tersebut. Tetapi, ketika proses bisnis tersebut dibuat akan ada masalah di dalamnya salah satunya adalah apakah proses bisnis yang dibuat mengalami duplikasi atau tidak.

Untuk mengatasi masalah di atas dapat dilakukan pengecekan kemiripan proses bisnis. Pengecekan ini juga dapat dilakukan apabila ada penggabungan dua organisasi. Ada beberapa teknik untuk mencari kemiripan proses bisnis ini berdasarkan *label similarity*, *structural similarity*, dan *behavioral similarity*. Dalam *label similarity* terdapat beberapa metode yang dapat dilakukan yaitu dengan *syntactic similarity* dan *semantic similarity*. Hasil dari *similarity* ini adalah antara 0 dan 1, dimana 0 menunjukkan apabila tidak ada kemiripan diantara proses bisnis ini dan 1 menunjukkan adanya kemiripan pada proses bisnis tersebut[1].

Dari permasalahan di atas, penelitian berfokus pada pengecekan kemiripan proses bisnis menggunakan teknik *label similarity* berdasarkan *syntactic* yang hanya berfokus pada pengecekan label. Pengecekan sintaks ini digunakan untuk mencari *edit distance*, dimana pencarian *edit distance* ini berguna untuk mengetahui berapa kali *string* melakukan *insert*, *delete*, maupun *substitution*. Pengecekan berdasarkan sintaks tersebut berguna apabila terdapat kata yang hampir mirip atau memiliki arti yang sama tetapi memiliki letak *string* yang berbeda. Misalnya melakukan terjemahan bahasa ke bahasa lain. Contohnya 'verification' dan 'verifikasi' dikatakan berbeda dikarenakan memiliki sintaks yang berbeda. Maka dari itu, untuk mencari kemiripan pada dua buah kata yang dibandingkan tidak hanya berfokus pada *semantic* tetapi berdasarkan sintaks juga perlu diperhatikan.

2. Landasan Teori

2.1 Pengertian Proses Bisnis

Terdapat beberapa definisi proses bisnis sebagai berikut:

- Menurut El sawy (2001,p16) proses bisnis adalah kegiatan kerja yang saling terorganisir berdasarkan logika berurutan dan sumber daya yang terkait didalamnya dapat menghasilkan suatu nilai kepada pelanggan [2].
- Menurut Hammer dan Champy (1993, p35) proses bisnis merupakan sekumpulan aktifitas yang memerlukan satu atau lebih masukan (input) dan memberntuk suatu keluaran (output) yang memiliki nilai yang diinginkan oleh pelanggan [3].
- Menurut Indrajit (2002,p3) proses bisnis adalah sejumlah aktivitas yang mengubah sejumlah input menjadi sejumlah output (barang dan jasa) untuk orang lain atau proses yang menggunakan orang dan alat [4].
- Menurut Hollander, Denna dan Cherrigton (2000,p4) proses bisnis adalah suatu urutan kegiatan yang menyelesaikan suatu tujuan bisnis [4].

Dari beberapa sumber yang telah didapatkan, dapat disimpulkan bahwa proses bisnis adalah suatu aktivitas yang mempunyai input dan berkaitan satu sama lain untuk menghasilkan sebuah output (barang atau jasa) yang dapat menunjang sebuah organisasi untuk mencapai visi dan misi dari organisasi tersebut.

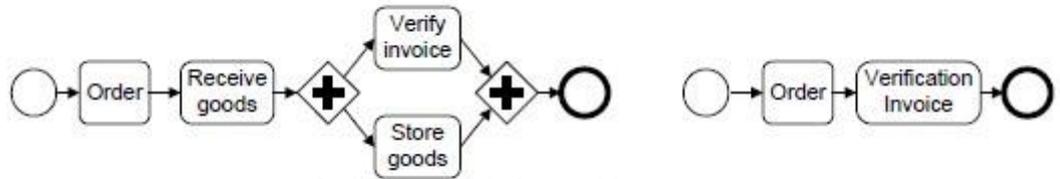
2.2 Similarity Proses Bisnis[5]

Kemiripan proses bisnis adalah adanya kesamaan proses bisnis dari organisasi. Untuk melakukan pencarian kemiripan dapat dilihat berdasarkan *label*, struktural, dan juga perilaku proses bisnis tersebut. Biasanya pengecekan kemiripan ini digunakan untuk kasus dimana terjadi penggabungan dua organisasi yang akan menyamakan visi dan misi mereka untuk membuat suatu model proses bisnis yang baru. Proses bisnis yang baru tidak perlu dibuat ulang, dengan melakukan pengecekan kemiripan proses bisnis ini kita dapat melakukan adaptasi ulang dari proses bisnis yang lama dengan cara melakukan *redesign* proses bisnis sesuai dengan kemiripan yang didapatkan dari hasil pencarian tersebut.

2.2.1 Label similarity[5]

Salah satu cara mencari kemiripan anatan hubungan model proses bisnis adalah dengan terlebih dahulu menghitung *alignment* antara model yaitu hubungan elemen-

elemen dalam satu model dengan model yang lainnya. Sebagai contoh gambar 2.1 yang dimodelkan menggunakan *Business Process Modelling and Notation*(BPMN).



Gambar 1 Sample pair of process [5]

Pada gambar 2.1, kegiatan yang sama yaitu “*verify invoice*” dengan “*verification invoice*”. Dilihat dari kesamaan kegiatan tersebut maka dapat didefinisikan sebagai perbandingan antara ukuran keselarasan, dan ukuran model proses. Sebagai contoh bias ditentukan ukuran kesamaan dengan formula sebagai berikut $\frac{2 \times |A \cap B|}{|A| + |B|}$ dimana A adalah kesamaan dan |A| dan |B| menunjukkan perbandingan jumlah elemen model di dalam P dan P’. Perhatikan bahwa “*Verify Invoice*” tidak sama dengan “*Verification Invoice*” maka harus memberikan bobot yang sama dengan kesamaan antar dua label ini. Ukuran kesamaan yang dihasilkan adalah pada persamaan 2.1:

$$S_{ij} = \frac{2 \times |A \cap B|}{|A| + |B|} \dots\dots\dots (2.1)$$

Dimana S_{ij} adalah keselarasan antara hubungan pada elemen proses. Kesamaan antara elemen model dapat dihitung dari label menggunakan ukuran kesamaan sintaks, kesamaan semantik, atau kombinasi dari keduanya. Perhitungan kesamaan ini berdasarkan sinonim kata dan hubungan semantik antara keduanya. Ada 3 teknik yang dilakukan untuk mencari kemiripan berdasarkan label, yaitu *semantic* dan *syntactic*.

2.2.2 Structural similarity[5]

Karena model proses adalah sebuah grafik, maka dapat menggunakan pencocokan grafik sebagai dasar untuk menentukan langkah-langkah kesamaan. Secara khusus, diberikan 2 model proses bisnis. Dapat didefinisikan kesamaan sebagai kebalikan dari jarak *graph-edit* mereka. Jarak *graph-edit* pada 2 graph adalah pencarian angka minimum pada operasi (*insertion, deletios, substitutions*) diperlukannya untuk perubahan grafik menuju ke yang lain. Metode yang digunakan dalam pencarian jarak minimum pada graph menggunakan A* Algorithm, *Heuristic Search*, dan *Similarity Flooding*.

2.2.3 Behavioral similarity[5]

Model proses adalah bagian dari pendefinisian perilaku sebuah organisasi. Kemungkinan lain untuk pengukuran kesamaan mereka adalah menggunakan *behavioral semantic*. *Behavioral* ini merupakan gabungan dari pencarian label berdasarkan semantik dan juga secara struktural.

2.3 Syntatic Similarity[1]

Kemiripan berdasarkan sintaks yaitu pencarian yang dapat ditemukan berdasarkan *edit distance* dari dua buah string yang dibandingkan. *Edit distance* itu sendiri memiliki arti yaitu perhitungan yang dilakukan apabila terjadi *delete, insert, ataupun subsitusi* yang terjadi pada satu *string* ke *string* lainnya.

Adapun persamaan untuk melakukan perhitungan *syntatic similarity*, yaitu:

$$S_{ij} = 1 - \frac{\text{edit distance}(s_i, s_j)}{\max(|s_i|, |s_j|)} \dots\dots\dots (2.2)$$

Untuk mencari kemiripan berdasarkan sintaks didapatkan dari hasil dikurangi pembagian *edit distance* dan *max*. Dimana seperti yang telah dijelaskan sebelumnya *edit distance* ini diihitung apabila *string* melakukan *insert, delete, atau subsitution* dari *string* satu ke *string* lainnya. *Max* didapatkan dari panjang karakter pada label 1 yang terdapat di node 1 dan label 2 yang terdapat di node 2. Setelah label pada node dibandingkan, yang memiliki panjang karakter paling besar maka akan menjadi nilai *max*.

2.4 Edit Distance[8]

Edit distance yang biasa disebut juga dengan *Leveinstein distance* adalah ukuran kesamaan antar dua string yang memiliki *distance* minimum, yang disebut sebagai string sumber (s) dan string sasaran (t). Jarak yang dimaksud adalah *delete*, *insert*, dan *substitutions* yang diperlukan untuk mengubah s menjadi t. Sebagai contoh, jika (s) adalah test dan (t) adalah test, maka *edit distance* (s,t) = 0, karena tidak ada transformasi yang diperlukan. String sudah identik. Contoh lain dari *edit distance* adalah kitten dan sitting. Kedua kata tersebut memiliki *edit distance* 3, yaitu :

1. Kitten → sitten (substitution of “s for “k”)
2. Sitten → sittin (substitution of “i” for “e”)
3. Sittin → sitting (insertion of “g” at the end)

Algoritma ini biasanya digunakan di:

- Pemeriksaan ejaan
- Pengenalan suara
- Analisis DNA
- Deteksi plagiarism

2.5 Business Process Modelling and Notation (BPMN)[6]

Business Process Modelling And Notation adalah standar yang sudah umum digunakan untuk pemodelan proses bisnis dan salah satu komponen yang paling penting untuk mencapai kesuksesan sebuah bisnis di bidang IT.

Beberapa alasan mengapa BPMN ini banyak digunakan oleh suatu organisasi untuk memodelkan proses bisnis mereka, diantaranya adalah:

- *Standar*: BPMN ini sudah ditetapkan melalui standar di seluruh dunia, misalnya UML. Standar ini didukung oleh banyak produk perangkat lunak jadi tidak perlu bergantung pada produk apapun.
- *Simplicity*: Prinsip dari BPMN sendiri adalah mudah digunakan sehingga pekerjaan akan cepat terselesaikan.
- *Power of expression*: Jika diperlukan, BPMN dapat menggambarkan dengan tepat bagaimana fungsi setiap proses.
- *Implementation in IT*: Fokus pengembangan BPMN ini untuk mendukung teknik implementasi pada proses (“*Process Automation*”).

2.6 Label Matching Similarity[1]

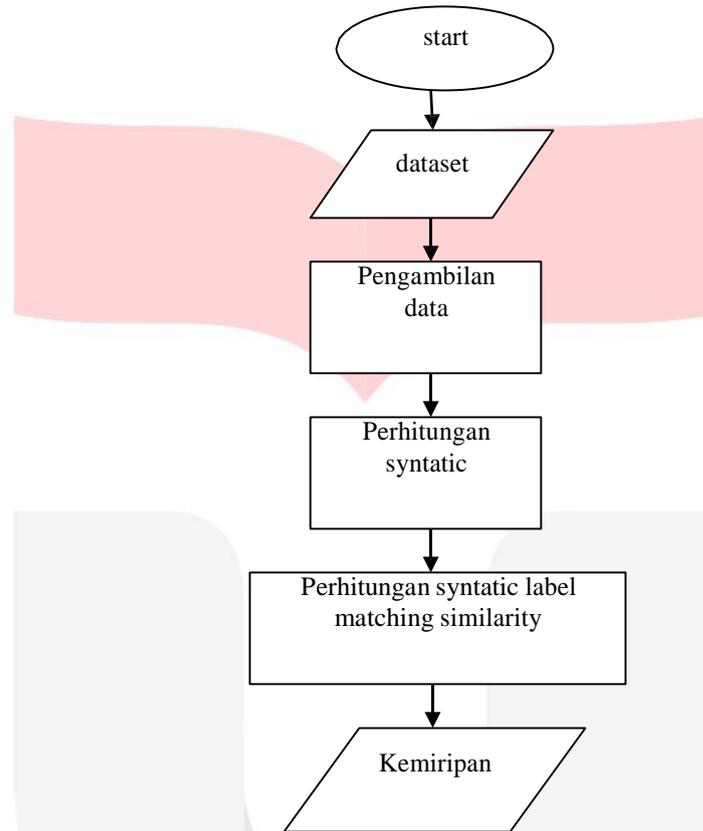
Label matching similarity ini di peroleh dari nilai mapping yang paling optimal antara node yang telah dibandingkan. Nilai *label matching similarity* ini adalah jumlah nilai *label* kemiripan pasangan node yang cocok. Untuk mendapatkan nilai antara 0 dan 1, dibagi jumlah tersebut dengan jumlah total node. Perhitungan persamaan *label matching similarity* adalah:

$$Sim(p_1, p_2) = \frac{\sum_{(i,j) \in M_{ij}} 1}{\sum_{i=1}^n 1} \dots\dots\dots (2.3)$$

Untuk melakukan *label matching similarity* ini dilakukan perhitungan kemiripan dengan cara *syntactic* atau *semantic* dan juga melakukan pencarian mapping paling optimal dari perhitungan *syntactic* atau *semantic* tersebut. Setelah itu untuk dimasukkan ke dalam perhitungan *label matching similarity*, harus diketahui terlebih dahulu perhitungan *syntactic* mana yang berada di atas *threshold*.

3. Gambaran Umum Sistem

Tahapan ini dilakukan perancangan untuk kerja sistem yang diimplementasikan. Sistem yang dibangun memiliki dua input yaitu proses bisnis pertama yang akan menjadi *query* atau pembandingan dan proses bisnis kedua yang akan menjadi *variant* yang dibandingkan dengan *query* yang struktur datanya adalah XML. File XML ini dihasilkan melalui penggambaran dengan menggunakan BPMN. File XML ini mengekstrak *task*, *parallel gateway*, dan *sequence flow*, dimana masing-masing *label* yang berada di dalam *node* akan diambil untuk dijadikan data dalam melakukan mapping. Dari mapping ini, dapat dilakukan perhitungan *syntactic similarity* dengan persamaan (2.2). Setelah dilakukan perhitungan tersebut, dapat dilakukan perhitungan tingkat kemiripan dengan menggunakan *label matching similarity*.



Gambar 2 Gambaran Umum Sistem

4. Hasil Pengujian

Berikut adalah hasil pengujian untuk perbandingan 2 model proses bisnis:

Tabel 1. Hasil pengujian

Perbandingan		Similarity
Query	Variant	
Reklame	STDP	80%
STDP	Reklame	
Reklame	UBPM	98,66%
UBPM	Reklame	
STDP	UBPM	72,72%
UBPM	STDP	

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat di simpulkan sebagai berikut:

- a. Cara pengecekan untuk mendapatkan tingkat kemiripan menggunakan *syntactic label matching similarity* dimana proses ini dilakukan berdasarkan perhitungan *syntactic similarity* terlebih dahulu lalu melakukan perhitungan *label matching similarity*.
- b. Tingkat kemiripan pada kedua proses bisnis memiliki nilai yang sama apabila dibolak-balik *query* menjadi *variant* ataupun sebaliknya. Tingkat kemiripan paling tinggi adalah perbandingan antara Reklame dan UBPM.
- c. Didapatkan hasil perbandingan antara Reklame dan STDP adalah 80%, Reklame dan UBPM 98,66% dan STDP dan UBPM adalah 72,72%.

5.2 Saran

Adapun saran yang diberikan penulis apabila penelitian ini dapat dikembangkan, yaitu:

- a. Dapat dilakukan dengan metode lain untuk mencari tingkat kemiripan berdasarkan *label* yaitu *semantic similarity*.
- b. Dapat dilakukan untuk mencari berdasarkan *structural similarity* agar dapat diketahui bagaimana bentuk proses bisnis yang baru ketika bergabung.
- c. Membuat *user interface* agar *user* lebih mudah untuk melakukan penginputan proses bisnis.

Daftar Pustaka

- [1] R. Djikman, M. Dumas, B. V. Dongen, R. Kaarik dan J. Mendling, "Similarity of Business Process Models: Metrics and Evaluation," 2009.
- [2] O. E. Sawy, Redesigning Enterprise Process for E-Business, McGraw Hill Higher Education, 2001, p. 16.
- [3] M. Hammer dan J. Champy, Reengineering the Corporation, 1993, p. 35.
- [4] P. Yessie dan S. , Peningkatan Kinerja Proses Pemenuhan Pesanan Pelanggan pada PT. Sri Indah Aluminium Extrusion, Jakarta, 2008.
- [5] L. B. R. D. M. Dumas, "Similarity Search of Business Process Models," pp. 23-28, 2009.
- [6] [Online]. Available: <http://www.boss.or.id>. [Diakses 16 January 2015].
- [7] [Online]. Available: <http://www.bpmn.info>. [Diakses 15 January 2015].
- [8] [Online]M. P. S. Michael Gilleland, "Leveinstein Distance, in Three Flavors,".[Diakses 2 Februari 2015].
- [9] S. Niila, "Analisis *Structural Similarity* Pada Model Proses Bisnis Menggunakan *Selective Reduce* Berbasis Petri Net," 2014
- [10] I. Leveinstein, "Binary code capable of correcting deletions, insertions, dan reversals," *Cybernetics and Control Theory*, vol. 10, no. 8, p. 707-710, 1980.