

PENGELOMPOKAN MODEL PROSES BERDASARKAN MATRIK SIMILARITAS DENGAN PENDEKATAN SEMANTIK

Bayu Priyambadha¹⁾ dan Riyanarto Sarno²⁾

¹⁾ Program Studi Teknik Informatika, Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya

²⁾ Program Studi Teknik Informatika, Fakultas Teknologi Informasi, ITS

Email : bayu_priyambadha@ub.ac.id¹⁾, riyanarto@if.its.ac.id²⁾

ABSTRAK

Saat ini, pemanfaatan sebuah sistem informasi berorientasi proses pada sebuah organisasi sangat marak dilakukan. Sistem informasi berorientasi proses bertujuan untuk meningkatkan kinerja sebuah organisasi. Dalam sebuah organisasi berskala besar, model proses yang digunakan untuk mendukung bisnis tidak berjumlah sedikit, melainkan dapat mencapai angka ratusan bahkan hingga ribuan. Repositori model proses adalah sebuah media untuk menyimpan model proses dalam sebuah organisasi. Terdapat permasalahan dalam pengelolaan repositori model proses, antara lain proses perhitungan kesamaan model proses yang masih menggunakan pendekatan kesamaan sintaktik. Pendekatan tersebut membuat proses pengelompokan model proses menjadi kurang optimal. Untuk menjawab permasalahan tersebut, pada penelitian ini dilakukan mekanisme pengelompokan model proses berdasarkan kedekatan derajat kesamaan yang dimiliki tiap model proses. Penghitungan derajat kesamaan dilakukan berdasarkan beberapa matrik kesamaan, antara lain kesamaan titik (node), kesamaan struktur, dan kesamaan perilaku (behavior). Serta perhitungan derajat kesamaan dilakukan dengan menggunakan metode kesamaan arti (semantik). Penggunaan metode kesamaan arti dapat meningkatkan nilai compactness pada kelompok yang dihasilkan dari proses clustering.

Kata Kunci: matrik kesamaan, pencarian (searching), pengelompokan (clustering), repositori model proses, semantik

ABSTRACT

Currently, the use of a process-oriented information system in an organization is very commonplace. Process-oriented information system aims to improve the performance of an organization. In a large-scale organization, process models are used to support the business does not amount to little, but can reach hundreds or even thousands. Repository model of the process is a medium for storing models of processes in an organization. There are problems in the management repository process models, among others the similarity calculation process models are still using syntactic similarity approach. Such an approach makes the process of grouping the process model to less than optimal. To answer these problems, in this research grouping mechanism process models based on the proximity degree of similarity that was owned by the process model. Calculation of the degree of similarity is based on some similarity matrix, among other similarities point (node), the similarity of structure, and the similarity of behavior. As well as the degree of similarity calculation is done using the similarity meaning (semantics). The use of semantics similarity method in can increases the value of compactness in the group resulting from the process of clustering.

Keywords: grouping (clustering), repository process model, search (searching), semantics, similarity matrix

I. PENDAHULUAN

PEMANFAATAN sistem informasi berorientasi proses di era saat ini sangat pesat. Sistem informasi berorientasi proses lebih sering dikenal sebagai PAIS (*Process Aware Information System*) [1]. Pengelolaan proses bisnis dalam sebuah sistem informasi yang berjalan pada suatu organisasi merupakan lingkup dari PAIS. PAIS sangat berperan penting dalam BPM (*Business Process Management*) [1].

Dalam sebuah organisasi yang berskala besar, model proses bisnis yang dijalankan tidak berjumlah sedikit. Model proses bisnis dirancang untuk mendefinisikan urutan proses untuk pencapaian suatu *goal* atau tujuan bisnis. Jumlah model proses bisnis dapat mencapai ratusan bahkan hingga ribuan model [2]. Kumpulan model proses disimpan di dalam sebuah tempat yang dinamakan repositori proses model. Jumlah yang besar merupakan sebuah tantangan bagi para peneliti untuk meneliti teknologi dalam mengoptimalkan repositori proses bisnis. Salah satu hal yang sering dibahas oleh beberapa peneliti adalah proses pencarian model proses bisnis dalam repositori [3][4][5][6]. Namun demikian, masih banyak diungkapkan kekurangan-kekurangan dari sistem pencarian yang ada. Antara lain, penelitian yang dilakukan oleh Dijkman et al [5], dimana diungkapkan suatu permasalahan yang muncul dari pencarian model proses dalam repositori adalah waktu pencarian yang lama. Dengan jumlah model proses yang sangat besar, proses pencarian yang dilakukan akan sangat lambat. Proses pencarian dilakukan dengan membandingkan model kueri dengan model dalam repositori. Proses pembandingannya dilakukan satu demi satu antara model kueri dan kelompok model. Sehingga proses pembandingan memerlukan usaha yang sangat besar. Yan et al. [6] memberikan jawaban atas permasalahan yang ada pada penelitian Dijkman et al [5] dengan memberikan

index pada repositori proses bisnis yang dibangun. Hasil yang diungkapkan oleh Yan et al. [6] menunjukkan adanya percepatan proses pencarian model proses bisnis.

Proses pencocokan model proses dalam repositori dapat dilakukan dengan membandingkan tiap-tiap model. Pencocokan model proses dapat berdasarkan pada matrik-matrik kesamaan model. Pada penelitian Sarno et. al. [10], mendefinisikan suatu cara dalam pengelompokan model proses dengan mempertimbangkan kesamaan struktural dan perilaku. Pendekatan yang dilakukan oleh Sarno adalah hanya mempertimbangkan kesamaan sintak pada label proses. Dijkman et al. [5], mendefinisikan terdapat 3 matrik pencocokan model proses bisnis, yaitu *node based*, *structural based*, dan *behavioral based similarity*. Ketiganya dinyatakan sebagai matrik yang dapat mengindikasikan derajat kesamaan antar model.

Pada penelitian ini, diterapkan sebuah metode dalam proses pengelompokan model proses. Proses pengelompokan dengan mempertimbangkan kesamaan model berdasarkan 3 matrik, yaitu kesamaan *node*, struktur (*structure*) dan perilaku (*behavior*). Dalam perhitungan kesamaan label pada model proses dipertimbangkan kesamaan semantik. Kesamaan semantik dilakukan dengan mempertimbangkan hubungan sinonim maupun hipernim dari setiap kata dalam label proses. Perhitungan kedekatan arti kata dalam sebuah label proses digunakan library Wordnet yang merupakan database kata-kata dan relasi semantiknya [12]. Data yang menjadi masukan pada sistem yang diajukan berupa file model dengan ekstensi XPDL (*XML Data Processing Language*) dan PNML (*Petri Net Markup Language*).

Penulisan makalah ini dibagi menjadi tujuh (7) bagian, bagian pertama adalah pendahuluan yang menjelaskan tentang latar belakang penelitian. Bagian kedua menjelaskan tentang landasan teori. Bagian ketiga menjelaskan metode yang digunakan atau urutan pengerjaan penelitian. Bagian keempat adalah proses implementasi metode. Bagian kelima analisa hasil percobaan yang akan disimpulkan pada bagian keenam. Bagian terakhir (ketujuh) adalah daftar pustaka.

II. LANDASAN TEORI

Terdapat beberapa hal yang melandasi penelitian ini. Dasar-dasar teori yang melandas penelitian ini antara lain adalah pemahaman tentang repositori model proses bisnis, pengelompokan model dan metode pencarian model. Masing-masing dasar teori akan dijelaskan pada sub bagian 2.1, 2.2 dan 2.3.

2.1 Repositori Model Proses Bisnis

Repositori model proses bisnis adalah sebuah media untuk menyimpan berbagai macam model bisnis proses dalam sebuah organisasi [2]. Model proses bisnis yang disimpan dalam repositori sangat beragam dan banyak. Jumlah model dalam repositori tergantung pada organisasi yang menggunakan dan mengelola. Semakin besar organisasi, semakin besar proses bisnis yang dijalankan, dan semakin besar juga jumlah model proses bisnis yang disimpan dalam repositori.

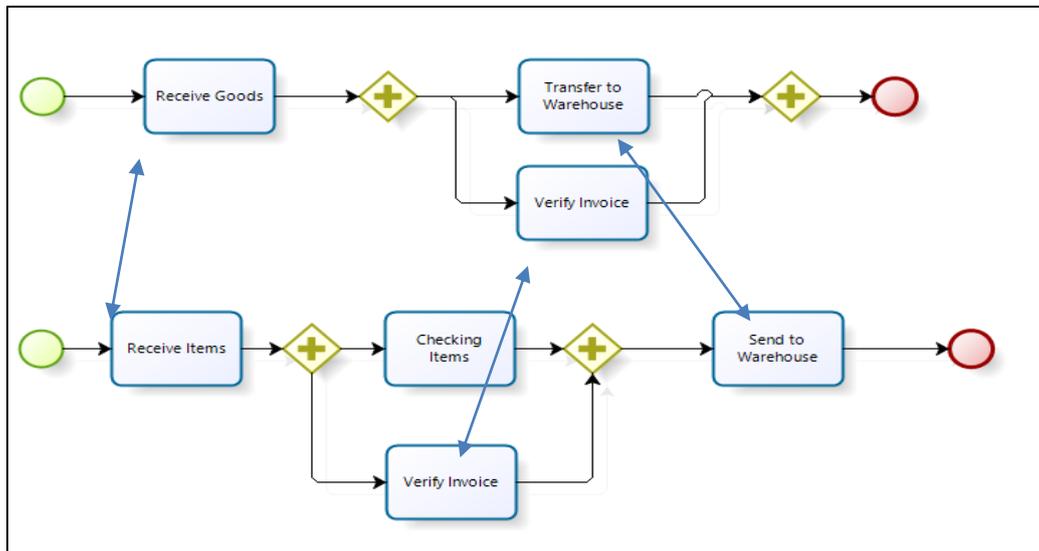
Repositori model proses bisnis membutuhkan sebuah perhatian yang lebih. Mengelola model proses dengan jumlah yang sangat banyak bukanlah hal yang mudah. Pengelolaan repositori model mempengaruhi kinerja sebuah organisasi. Pengelolaan yang baik akan memberikan percepatan pada proses pencarian atau pencocokan model proses bisnis [2]. Beberapa informasi dapat ditambahkan pada repositori untuk meningkatkan kinerja proses pencarian dan pencocokan model bisnis proses [2].

2.2 Pengelompokan Model (*Model Clustering*)

Salah satu tujuan pengelolaan repositori bisnis proses adalah untuk mencapai proses pencarian atau pencocokan yang cepat dan tepat. Untuk meningkatkan proses pencarian atau pencocokan dapat dilakukan dengan metode pengelompokan (*clustering*) [7]. Pengelompokan pada model proses bisnis dilakukan dengan membuat kelompok-kelompok model berdasarkan derajat kesamaan yang ada pada model proses [7]. Jung et al. [7] melakukan pengelompokan berdasarkan matrik kesamaan struktural.

2.3 Matrik Kesamaan Model Proses

Terdapat beberapa matrik kesamaan yang dapat digunakan dalam melakukan pengelompokan model proses. Jung et al. [7] hanya menggunakan satu matrik yaitu kesamaan struktural model proses. Dijkman et al. [5] mendefinisikan beberapa matrik kesamaan model proses. Matrik-matrik tersebut dijelaskan pada sub bagian selanjutnya.



Gambar 1. Contoh Perbandingan Dua Model Proses

Gambar 1 menjelaskan contoh dua model proses tentang penerimaan barang. Dua model tersebut dilihat dari tujuan akhir (hasil akhir) memiliki tujuan yang sama. Masing-masing mengelola penerimaan barang sampai proses penyimpanan. Pada bagian selanjutnya akan menggunakan contoh model proses pada Gambar 1 untuk melakukan simulasi perhitungan nilai kesamaan.

2.3.1 Matrik Kesamaan Node

Matrik kesamaan *node*, adalah matrik yang dilihat dari atribut yang ada pada tiap node proses. Proses perhitungan kesamaan dilakukan dengan membandingkan *node* yang terdapat pada masing-masing proses model. Kesamaan *node* didapatkan dengan menjumlahkan nilai kesamaan label *node* yang relevan dan dibagi dengan total *node* yang ada pada model proses. Rumus perhitungan kesamaan node [5] pada model proses dapat diformulasikan seperti berikut.

$$simnm(B1, B2) = \frac{2 \times \sum_{(n,m) \in M_{Sim}^{Opt}} Sim(n,m)}{|\{n|n \in N_1\}| + |\{n|n \in N_2\}|} \quad (1)$$

Perhitungan nilai kesamaan *node*, dilakukan dengan memilih pasangan *node* antar model proses yang paling optimal. Node yang optimal dicapai dengan mempertimbangkan ambang batas (*threshold*) tingkat kesamaan yang sudah didefinisikan terlebih dahulu.

Dari contoh kasus yang tergambar pada Gambar 1, perhitungan kesamaan node dapat kita lakukan. Dalam kasus ini, ambang batas kita asumsikan bernilai 0.5. Nilai ambang batas dapat kita gunakan untuk menemukan pasangan proses yang optimal. Tiga pasang proses dinyatakan sebagai pasangan yang memiliki kemiripan sintak, yaitu “Receive Goods” dan “Receive Items”, “Verify Invoice”, dan “Transfer to Warehouse” dan “Sent to Warehouse”. Jumlah seluruh node pada kedua model proses tersebut yang tidak memiliki pasangan adalah 9 proses. Dari perhitungan tersebut, nilai kesamaan node adalah $\frac{2 \times (0.69 + 1 + 0.66)}{9} = 0.52$.

2.3.2 Matrik Kesamaan Struktur

Matrik kesamaan struktur, adalah matrik yang dilihat dari kesamaan bentuk struktur model proses. Nilai kesamaan struktur dilihat menggunakan metode *graph-edit distance* [5]. Proses tersebut adalah penerapan prinsip metode *edit distance* pada *graph*. Seperti halnya *edit distance*, *graph-edit distance* juga mempertimbangkan *substitution*, *inserted* dan *deleted*. Ketiga hal tersebut diterapkan pada *node* dan *edge* pada model proses.

Seperti pada perhitungan kesamaan *node*, pada proses ini dihitung juga perbandingan antar *node* yang diikuti dengan perhitungan *edit distance* pada *graph*. Formula perhitungan *graph-edit distance* [5] digambarkan pada rumus berikut ini.

$$simged(B1, B2) = 1 - avg(snv, sev, sbv) \quad (2)$$

Dimana :

$$\begin{aligned}
 snv &= \frac{sn}{N_1 + N_2}, \\
 sev &= \frac{se}{E_1 + E_2}, \\
 sbv &= \frac{2 \times \sum_{(n,m) \in M} 1 - Sim(n,m)}{N_1 + N_2 - sn}
 \end{aligned} \tag{3}$$

sn adalah jumlah inserted atau deleted node. se adalah jumlah inserted atau deleted edge. N_1 dan N_2 merupakan jumlah *node* pada setiap *graph*. E_1 dan E_2 adalah jumlah *edge* dalam setiap *graph*.

Nilai kesamaan struktur kasus Gambar 1 dapat dihitung dengan proses sebagai berikut. Nilai snv adalah nilai node inserted atau deleted dari kedua model dibagi dengan jumlah seluruh node, sehingga didapatkan nilai $snv = \frac{9}{15} = 0.6$. Nilai sev didapatkan dengan membagi nilai edge inserted atau deleted dengan jumlah seluruh edge, sehingga didapatkan nilai $sev = \frac{11}{15} = 0.73$. Dan nilai sbv adalah sebagai berikut $sbv = \frac{2(1-0.69+1-1+1-0.66)}{15-9} = 0.21$. Dari ketiga nilai tersebut maka, nilai similaritas struktur adalah $simged = 1 - 0.514 = 0.486$.

2.3.3 Matrik Kesamaan Behavior

Matrik kesamaan *behavior* (perilaku), adalah matrik yang dilihat dari hubungan kausal antara label proses dalam sebuah model proses. Hubungan kausal dapat diartikan sebagai urutan proses yang ada dalam model proses.

Perhitungan kesamaan model proses dengan melihat kesamaan perilaku (*behavior*) dilakukan dengan menghitung kesamaan urutan proses dalam model proses. Selanjutnya, jumlah proses yang berurutan dalam dua model proses akan dibagi dengan jumlah total urutan model proses. Mendling et al [8], menyatakan kesamaan perilaku dapat dihitung dengan metode *Naïve Behavioral Similarity* (NBS). NBS dapat dihitung dengan membandingkan urutan proses antara dua model. NBS dapat dirumuskan sebagai berikut :

$$simbehav(B1, B2) = \frac{NumPairSq(NS_1, NS_2)}{NS_1 + NS_2} \tag{4}$$

Dimana, NS_1 dan NS_2 adalah urutan proses yang ada pada model pertama dan kedua, sehingga apabila kita terapkan pada kasus Gambar 1 maka nilai similaritas perilaku kedua model tersebut dapat dihitung. Pasangan aktifitas yang ada antara model 1 dan model 2 adalah 5 pasang (beririsan). 1 pasang merupakan pasangan yang sama antara kedua model. Nilai similaritas perilaku adalah $simbehav = \frac{1}{5} = 0.2$.

2.4 Kesamaan Semantik

Kesamaan secara semantik dapat diartikan kesamaan berdasarkan pada ekivalensi antara dua kata. Ekivalensi antar kata dapat berupa nilai derajat kesamaan arti (sinonim) antara keduanya [5].

l_1 dan l_2 adalah label dari sebuah proses yang akan dibandingkan. Secara logika, l_1 dan l_2 memiliki kemungkinan untuk mempunyai nilai kesamaan dilihat dari sinonimnya berdasarkan kamus bahasa. Salah satu referensi atau kamus dalam bentuk digital adalah WordNet [9]. Fellbaum dalam Agirre [9] menyatakan bahwa WordNet adalah database leksikal dalam bahasa inggris, yang berisi kelompok kata, kata kerja, kata sifat dan adverbial yang dikelompokkan menjadi sekelompok sinonim kata (*synsets*). WordNet dapat digunakan sebagai alat untuk mengukur derajat keterkaitan arti antara dua kata. Dalam pendefinisian label pada proses, diasumsikan kesamaan arti dapat disamakan dengan kesamaan maksud atau kesamaan aktifitas proses.

Dalam penelitian yang dilakukan oleh Dijkman [5], Dijkman merumuskan sebuah rumusan untuk menghitung kesamaan antara dua label dengan mempertimbangkan kesamaan arti (sinonim). Kesamaan semantik dapat dihitung dengan menggunakan rumus berikut ini :

$$\frac{2 \cdot w_i \cdot |w_1 \cap w_2| + w_s \cdot (|s(w_1, w_2)| + |s(w_2, w_1)|)}{|w_1| + |w_2|} \quad (5)$$

dimana, w_1 dan w_2 adalah kumpulan kata dari setiap label yang akan dibandingkan. $s(w_1, w_2)$ atau $s(w_2, w_1)$ adalah jumlah kata yang memiliki hubungan sinonim antara dua label. w_i dan w_s adalah bobot yang didefinisikan untuk kata yang sama dan kata yang memiliki kesamaan arti (sinonim). Dijkman [5], mendefinisikan nilai $w_i = 1$ dan $w_s = 0.75$.

2.5 WordNet

WordNet adalah database yang menghubungkan kata benda, kata kerja, kata sifat, dan kata keterangan dalam bahasa inggris sehingga membentuk sebuah relasi sinonim [12]. Relasi tersebut dapat digunakan untuk menentukan makna kata dari kata-kata yang terhubung. *WordNet* merupakan database relasi antar kata. Beberapa library pemrograman telah dibangun untuk mempermudah pencarian kata maupun untuk menghitung kedekatan makna antara dua kata. *RitaWordnet* adalah salah satu librari yang dapat digunakan untuk menghitung kedekatan makna antara dua kata. Librari tersebut dibangun menggunakan bahasa Java.

III. METODOLOGI

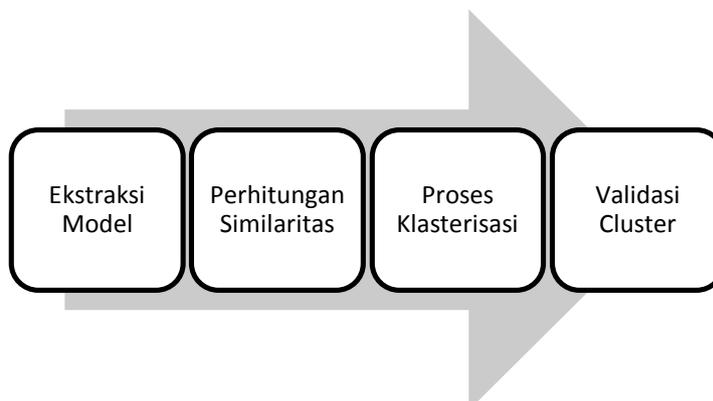
Penelitian ini dilakukan dengan beberapa tahapan. Tahapan-tahapan tersebut dilakukan berurutan yang dimulai dari proses ekstraksi model dari file XPDL atau PNML menjadi model graf, diikuti dengan proses perhitungan kesamaan berdasarkan matrik kesamaan node, struktur dan behavior (perilaku), lalu proses pengelompokan model proses berdasarkan nilai kedekatannya kesamaannya. Tahapan terakhir adalah proses validasi klaster yang terbentuk. Urutan tahapan penelitian ini digambarkan pada Gambar 2.

Dalam proses perhitungan similaritas antara dua model proses, akan digunakan tiga matrik yang didefinisikan oleh Dijkman [5]. Matrik-matrik itu adalah kesamaan *node*, kesamaan struktur dan kesamaan *behavior*. Dalam penelitian ini tidak diberikan pembobotan pada masing-masing matrik, semua matrik memiliki bobot yang sama.

Dalam penelitian Sarno et. al. [10] dijelaskan bahwa pemberian bobot pada masing-masing matrik tidak memberikan dampak yang signifikan pada hasil perhitungan. Proses perhitungan similaritas berdasarkan kesamaan arti (sinonim) ditambahkan untuk meningkatkan keakuratan dalam perbandingan kesamaan antara dua model proses. Dalam proses perbandingan dua model proses menggunakan ambang batas sebesar 0.5. Penentuan ambang batas sebesar 0.5 berdasarkan bahwa angka 0.5 adalah angka yang berada pada posisi tengah dari rentang angka 0 hingga 1. Proses perhitungan kedekatan sinonim dilakukan dengan menerapkan rumus (5) yang didefinisikan oleh Dijkman [5], dengan mengganti fungsi $s(w_1, w_2)$ dan $s(w_2, w_1)$ dengan librari *WordNet* agar proses perhitungan dapat dilakukan secara otomatis.

Proses selanjutnya adalah proses klasterisasi atau pengelompokan. Proses klasterisasi atau pengelompokan model dilakukan dengan tiga ambang batas yaitu 0.5, 0.6, dan 0.7 [10]. Ambang batas tersebut disamakan dengan nilai yang digunakan pada penelitian Sarno et. al. [10]. Pembagian menjadi tiga kelompok percobaan dengan maksud untuk mengetahui perbedaan hasil pembagian dengan ambang batas yang berbeda.

Proses validasi adalah proses yang paling akhir. Tujuan dari proses validasi adalah untuk mengetahui seberapa valid pembagian kelompok yang sudah dihasilkan. Proses validasi dilakukan dengan menghitung tingkat kohesi atau kedekatan antar model dalam satu kelompok, serta antar kelompok. Metode yang digunakan untuk melakukan validasi adalah Silhouette index [11]. Metode ini juga digunakan pada penelitian Sarno [10].



Gambar 2. Tahapan Proses Penelitian

TABEL I
HASIL PENGELOMPOKAN DENGAN AMBANG BATAS 0.7

Ambang Batas 0.7											
Kelompok (MTO)											
1	2	3	4	5	6	7	8	9	10	11	12
01	05	06	09	10	11	15	16	17	18	19	27
02		07			12					20	28
03		08			13					21	
04					14					22	
										23	
										24	
										25	
										26	

TABEL II
HASIL PENGELOMPOKAN DENGAN AMBANG BATAS 0.6

Ambang Batas 0.6				
Kelompok				
1	2	3	4	5
MTO01	MTO15	MTO18	MTO19	MTO27
MTO02	MTO16		MTO20	MTO28
MTO03			MTO21	
MTO04			MTO22	
MTO05			MTO23	
MTO06			MTO24	
MTO07			MTO25	
MTO08			MTO26	
MTO09				
MTO10				
MTO11				
MTO12				
MTO13				
MTO14				
MTO17				

Aplikasi pengelompokan bisnis proses dibangun menggunakan bahasa Java. Pemanfaatan database *WordNet* menggunakan sebuah librari Java yaitu *RitaWordNet*. Librari tersebut dapat digunakan untuk menghitung kedekatan makna antara dua kata dengan mempertimbangkan kesamaan makna kata. *RitaWordNet* menggunakan kamus *WordNet* untuk menghitung kesamaan makna kata. Selain *RitaWordNet*, librari *JDOM* juga digunakan dalam pembangunan aplikasi ini. *JDOM* merupakan sebuah librari yang digunakan untuk membaca data dalam format XML.

IV. PERCOBAAN

Percobaan dilakukan dengan menggunakan data model proses. Model proses dibuat dengan menggunakan notasi *PetriNet* yang berekstensi *PNML*. Model proses yang digunakan berjumlah 28 (dua puluh delapan) model. Model didapatkan dari aplikasi ERP pada penelitian Sarno et. al. [10]. Proses perhitungan kesamaan model dilakukan dengan menggunakan pendekatan kesamaan semantik dalam tiga matrik kesamaan. Tiga matrik yang digunakan dalam proses perhitungan kesamaan adalah node, struktur dan *behavior*. Didalam fungsi perhitungan kesamaan disisipkan fungsi kesamaan semantik.

Alat bantu perhitungan dibangun dengan bahasa Java, dengan menerapkan rumusan perhitungan yang sudah dijelaskan pada bagian sebelumnya. Hasil perhitungan secara otomatis disimpan dalam file excel untuk proses pembagian kelompok model berdasarkan kesamaan tiga matrik. Pembagian kelompok model dilakukan dengan tiga ambang batas. Hasil pengelompokan ditampilkan pada Tabel I, II dan III.

TABEL III
HASIL PENGELOMPOKAN DENGAN AMBANG BATAS 0.5

Ambang Batas 0.5	
Kelompok	
1	2
MTO01	MTO18
MTO02	
MTO03	
MTO04	
MTO05	
MTO07	
MTO08	
MTO09	
MTO10	
MTO11	
MTO12	
MTO13	
MTO14	
MTO15	
MTO16	
MTO17	
MTO19	
MTO20	
MTO21	
MTO22	
MTO23	
MTO24	
MTO25	
MTO26	
MTO27	
MTO28	

TABEL IV
PERBANDINGAN RATA-RATA NILAI SILHOUETTE INDEX

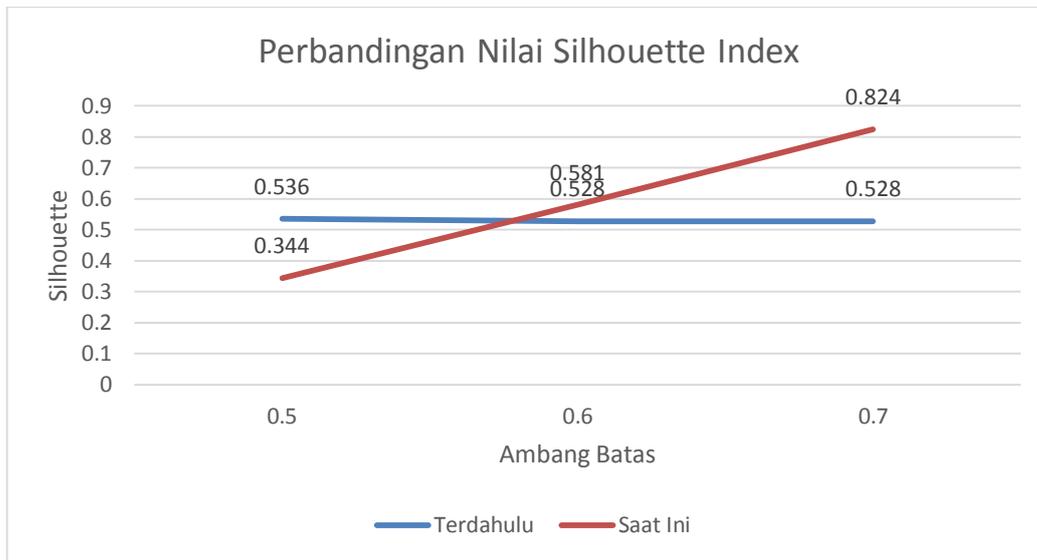
Threshold	Method	
	Previous	Proposed
0.5	0.536	0.344
0.6	0.528	0.581
0.7	0.528	0.824

V. VALIDASI DAN ANALISA HASIL

Dalam analisa dan validasi hasil dilakukan proses validasi hasil proses pengelompokan dengan menggunakan metode Silhouette index. Pada tahap ini nilai validasi akan dibandingkan dengan hasil yang dijelaskan pada penelitian Sarno et. al. [10]. Dari hasil perbandingan tadi akan ditarik kesimpulan bagaimana nilai index yang dihasilkan dari metode yang dipaparkan Sarno dengan metode yang diajukan. Nilai Silhouette index yang diambil dari penelitian Sarno adalah nilai rata-rata dari setiap percobaan dengan bobot berbeda tiap matrik.

Perbandingan hasil perhitungan Silhouette index pada dataset yang sama dijelaskan pada Tabel IV. Dari data hasil perhitungan pada Tabel IV, digambarkan dalam bentuk grafik pada Gambar 3.

Hasil perbandingan pada Gambar 3 menunjukkan nilai yang berbeda dari tiap ambang batas yang digunakan. Pada penelitian terdahulu [10], nilai Silhouette dari setiap ambang batas memiliki nilai yang hampir sama, bahkan mendekati sama. Tidak ada peningkatan nilai ketika nilai ambang batasnya meningkat dan kelompok yang



Gambar 3. Grafik Perbandingan Nilai Silhouette Index

dihasilkan berbeda. Pada penelitian ini, nilai Silhouette yang dihasilkan beragam. Nilai Silhouette semakin meningkat seiring dengan meningkatnya nilai ambang batas dan meningkatnya jumlah kelompok yang terbentuk. Secara logika sederhana apabila nilai ambang batas meningkat maka toleransi akan berkurang sehingga jumlah kelompok yang dihasilkan meningkat. Hal tersebut mengindikasikan juga meningkatnya nilai Silhouette *index* pada masing-masing kelompok karena terjadi peningkatan kohesi atau *compactness* (kepadatan) pada suatu kelompok.

VI. KESIMPULAN

Proses pengelompokan pada model proses merupakan hal yang sangat diperlukan untuk meningkatkan proses pencarian pada sebuah repositori. Penggunaan tiga matrik kesamaan dapat meningkatkan hasil pengelompokan. Kelompok yang dihasilkan lebih optimal karena semakin banyak parameter yang dipertimbangkan akan semakin baik hasilnya. Penggunaan kesamaan arti dalam proses perbandingan label proses menduduki posisi yang sangat vital dalam pencocokan proses model. Kesamaan proses model tidak hanya dilihat dari kesamaan sintak tapi kesamaan arti juga dipertimbangkan. Hal tersebut membuat hasil pengelompokan semakin optimal. Hal ini terbukti dari hasil perhitungan *compactness* dengan metode Silhouette *index* yang menunjukkan adanya peningkatan nilai ketika ambang batas dinaikkan. Ambang batas yang paling optimal adalah 0.7 karena pada posisi tersebut nilai Silhouette *index* mencapai 0.824.

Pada penelitian mendatang, akan ditambahkan matrik kesamaan yang lain. Matrik kesamaan kontekstual adalah matrik yang dirasa perlu dipertimbangkan. Nilai kesamaan sebuah proses model tidak hanya dilihat dari kesamaan label saja tapi dapat dilihat dari level yang lebih tinggi yaitu dari koteksnya. Hal lain yang akan ditambahkan untuk meningkatkan proses pengelompokan adalah dengan menambahkan orientasi perhitungan kesamaan. Sebagai perumpamaan, apabila kita mengukur jarak antara kota A dan B dimulai dari kota A maka hasilnya akan mungkin berbeda apabila kita memulai pengukuran dari kota B. Dengan permisalan tersebut, pada penelitian mendatang akan ditambahkan dari proses model mana nilai kesamaan itu akan dihitung.

DAFTAR PUSTAKA

- [1] Aalst, W.M.P.V.D. *Process-Aware Information Systems: Design, Enactment, and Analysis*
- [2] Dijkman, R., La Rosa, M., Reijers H.A. 2011. *Managing Large Collections of Business Process Models – Current Techniques and Challenges*. *Computers in Industry*, 63(2), pp. 91-97.
- [3] Dongen, B.F.v., Dijkman, R.M., Mendling, J. 2008. *Measuring Similarity between Business Process Models*. *CAiSE 2008 Montpellier*. Pp. 450-464.
- [4] Dijkman, R., Dumas, M., Garcia-Banuelos, L. 2009. *Graph Matching Algorithms for Business Process Model Similarity Search*. *7th International Conference, BPM 2009*. Pp. 48-63.
- [5] Dijkman, R., Dumas, M., Dongen, B.v., Kaarik, R., Mendling, J. 2010. *Similarity of Business Process Models : Metrics and Evaluation*. *Information System* 36. Pp. 498-516.
- [6] Yan, Z., Dijkman R., Grefen, P. 2010. *Fast Business Process Similarity Search with Feature-Based Similarity Estimation*. *OTM 2010, Part I, LNCS 6426*, pp. 60-77. Berlin : Springer-Verlag
- [7] Jung, J.Y., Bae, J., Liu L. 2008. *Hierarchical Clustering of Business Process Models*. *IEEE International Conference on Service Computing, IEEE Computer Society*. Pp. 613-616.
- [8] Mendling, J., Dongen, B.V., Aalst, W.V. 2007. *On the Degree of Behavioral Similarity between Business Process Models*. In proceeding of: 6. Workshop der Gesellschaft für Informatik e.V. (GI) und Treffen ihres Arbeitskreises

- [9] Agirre, E., Alfonseca, E., Hall, K., Kravalova, J., Pasca, M., Soroa, A. 2009. A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches. Association for Computational Linguistics. Pp. 19-27
- [10] Sarno, R., Ginardi, H., Pamungkas, E. W., & Sunaryono, D. 2013. Clustering of ERP Business Process Fragments, 2-7.
- [11] Rousseeuw, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65. doi:10.1016/0377-0427(87)90125-7
- [12] G. a. Miller, "WordNet: a lexical database for English," *Commun. ACM*, vol. 38, no. 11, pp. 39-41, 1995.