

## MEASURING BUSINESS PROCESS SIMILARITY USING PROBABILISTIC LATENT SEMANTIC ANALYSIS (PLSA) AND GREEDY GRAPH MATCHING

Fildzah Shabrina<sup>(1)</sup>, Riyanarto Sarno<sup>(2)</sup>

Program Studi Sistem Komputer, Fakultas Ilmu Komputer, Universitas Borobudur Jakarta  
Jl. Tarum Barat Kalimalang, Cipinang Melayu, Kec. Makasar, Kota Jakarta Timur  
e-mail : fildzah@borobudur.ac.id<sup>(1)\*</sup>

### Abstract

*The business process is a set of activities and tasks performed to achieve the goals of an organization. The business process model can be reused as a business process management effort into a repository. To solve the problem, it is necessary to measure the business process model that has similarity or similarity in terms of activity or process. From several business process models that have similarity can be identified as the main business process model, which has the primary function of the same activity. Business process model matching is the one of technique that can be used to identify, to measure the similarity of a set of business process models. The graph matching approach fit to identify the similarity of processes or activities in the business process model. The technique of matching the graph with Greedy graph matching shows similar results with an 89% precision value based on measuring the similarity of the graph building structure. Another approach in graph matching is a semantically or a text-based. Probabilistic Latent Semantic Analysis (PLSA) is one of the semantic approaches to measure the similarity of text in documents. PLSA measures the linkage of words in the document to identify any similarity of topics in the document. Measuring PLSA in business process matching analysis is by comparing text labels on each node in the business process. This research measures the similarity of business process models by combining two similarity analysis techniques based on semantics using PLSA and structural with Greedy. A graph matching technique by computing the semantics of each label on activities that are related to other activity labels. Structurally, connected activities are related to the same process or the same function. The result of this research is to know the effectiveness of business process which has activity relation.*

**Keywords : Business Process, BPMN, Graph Similarity, Probabilistic Latent Semantic Analysis (PLSA), Greedy Graph Matching**

Proses bisnis adalah serangkaian aktivitas dan tugas yang dilakukan untuk mencapai tujuan dari sebuah organisasi. Model proses bisnis dapat digunakan kembali sebagai upaya manajemen proses bisnis tersebut ke dalam sebuah repositori. Dalam repositori berisi ratusan hingga ribuan model proses bisnis dengan model yang sama maupun berbeda. Hingga dapat terjadinya duplikasi dan penumpukkan data. Untuk mengatasi permasalahan tersebut, perlunya dilakukan pengukuran terhadap model proses bisnis yang memiliki kesamaan atau kemiripan dalam hal aktivitas ataupun proses. Beberapa model proses bisnis yang memiliki kemiripan (similarity) dapat diidentifikasi sebagai model proses bisnis utama, yaitu memiliki fungsi dan aktivitas yang sama. Mencocokkan model proses bisnis merupakan salah satu teknik untuk mengidentifikasi, mengukur kemiripan dari kumpulan model proses bisnis. Pendekatan pencocokkan graf (graph matching) cocok untuk mengidentifikasi kemiripan proses atau aktivitas dalam model proses bisnis. Teknik mencocokkan graf dengan Greedy graph matching menghasilkan nilai presisi sebesar 89% berdasarkan pengukuran kemiripan struktur graf. Pendekatan lain dalam pencocokkan graf ialah secara semantik atau teks. Probabilistic Latent Semantic Analysis (PLSA) merupakan salah satu pendekatan semantik untuk menghitung kemiripan teks dalam dokumen. Perhitungan PLSA dalam analisis pencocokkan proses bisnis adalah dengan membandingkan label teks pada tiap node (label) proses bisnis. Penelitian ini mengukur kemiripan model proses bisnis dengan menggabungkan dua teknik analisis kemiripan berdasarkan semantik menggunakan PLSA dan struktural dengan Greedy. Teknik pencocokkan graf dengan menghitung semantik dari setiap label aktivitas yang saling memiliki keterkaitan atau hubungan. Secara struktural, beberapa aktivitas saling terhubung memiliki keterkaitan proses atau fungsi yang sama. Hasil penelitian ini adalah untuk mengetahui efektifitas dari proses bisnis yang memiliki keterkaitan aktivitas.

**Kata Kunci : Proses Bisnis, BPMN, Kemiripan Graf, Probabilistic Latent Semantic Analysis (PLSA), Greedy Graph Matching.**

## 1. PENDAHULUAN

Proses bisnis merupakan serangkaian aktivitas dan tugas yang dilakukan untuk mencapai tujuan dari sebuah organisasi. Proses bisnis dapat dikatakan sebagai alat atau instrumen untuk mengatur keseluruhan kegiatan serta untuk menghasilkan proses bisnis yang optimal dapat mengacu pada pendekatan Manajemen Proses Bisnis (Anugrah dan Sarno:2016). Faktor-faktor yang berkontribusi untuk menghasilkan proses bisnis yang optimal, baik secara langsung maupun tidak langsung, dapat ditinjau dari segi proses manajemen, proses operasional, dan pendukung proses bisnis tersebut. Model proses bisnis dapat digunakan kembali sebagai upaya manajemen proses bisnis tersebut ke dalam sebuah repositori atau kumpulan dari proses bisnis yang telah dibentuk (Djikman:2011). Repositori tersebut berisi ratusan hingga ribuan model proses bisnis. Manajemen proses dapat dilakukan untuk mengolah sekian banyak model proses bisnis. Salah satu contohnya adalah mengecek kesamaan atau kemiripan dari model proses bisnis (Djikman:2011) sebelum ditambahkan ke dalam repositori, hingga untuk menghindari duplikasi dan penumpukan data.

Untuk mengatasi permasalahan tersebut, maka dibutuhkan pengukuran terhadap model proses bisnis yang memiliki kesamaan atau kemiripan dalam hal aktivitas ataupun prosesnya. Dari beberapa model proses bisnis yang memiliki kemiripan (*similarity*) dapat diidentifikasi sebagai model proses bisnis utama, yang memiliki keutamaan fungsi, hingga aktivitas yang sama. Mencocokkan model proses bisnis (*matching process*) menjadi salah satu teknik yang dapat digunakan untuk mengidentifikasi, mengukur kemiripan dari kumpulan model proses bisnis (Kuss dan Stuckenschmidt: 2017). Tahap pertama dalam membandingkan model proses bisnis adalah dengan menentukan aktivitas mana yang memiliki keterkaitan, kesesuaian dalam aktivitas satu model proses bisnis dengan aktivitas di model lainnya (Weidlich, Djikman, dan Mendling). Langkah ini disebut pencocokkan model proses bisnis.

Menurut Becker dan Laue (2011), pendekatan pencocokkan graf (*graph matching*) cocok untuk mengidentifikasi kemiripan proses atau aktivitas dalam model proses bisnis. Sebuah graf memiliki *node* (simpul) dan *edge* (sisi). Dalam memetakan proses bisnis ke dalam sebuah graf ditunjukkan dengan *node* sebagai aktivitas dan *edge* sebagai proses atau fungsi yang terjadi antar satu aktivitas dengan aktivitas lain. Kemiripan model proses bisnis dideteksi dengan adanya kesesuaian proses dan fungsi yang sama pada tiap aktivitas.

Djikman (2009) mendeskripsikan tiga hal dalam menemukan kemiripan antara tiap model proses, diantaranya kemiripan teks, kemiripan struktur, dan kemiripan perilaku. Kemiripan teks muncul berdasarkan perbandingan bahasa yang digunakan, bisa dalam bentuk kalimat atau kata. Pada model proses bisnis ialah pada kalimat pendeskripsi aktivitas yang dilakukan. Kemiripan struktur ditinjau dari topologi atau struktur pembentuk, bila direpresentasikan dalam graf, maka struktur graf tersebut. Sedangkan kemiripan perilaku lebih mengacu kepada eksekusi proses semantik pada model proses.

Algoritma pencarian *Greedy* merupakan salah satu teknik yang dapat diterapkan dalam proses mencocokkan graf (Djikman dan Dumas: 2009). Teknik mencocokkan graf dengan *Greedy* (*Greedy graph matching*) menunjukkan hasil nilai presisi 89% berdasarkan pengukuran kemiripan struktur graf. Hasil ini menunjukkan bilamana 89% *Greedy* dengan tepat menemukan graf yang memiliki kemiripan sesuai dengan struktur perbandingan graf. Mencocokkan graf dengan *Greedy* menghasilkan nilai optimal dan waktu eksekusi dibandingkan dengan *A Star* (Djikman dan Dumas: 2009)., dengan menghitung label yang mirip atau sepadan, dan mengecek bilamana ada atau tidak *node* yang terhubung memiliki kecocokkan.

Pendekatan lain dalam pencocokkan graf ialah secara semantik atau berbasis teks. *Probabilistic Latent Semantic Analysis* (PLSA) merupakan salah satu pendekatan semantik untuk menghitung kemiripan teks dalam dokumen. Dengan menghubungkan dokumen dan kata sebagai kata kunci memodelkan variabel dalam PLSA (Sarno: 2016). PLSA menunjukkan hasil

lebih baik dalam proses pencocokkan dibandingkan dengan metode VSM dan LSA dengan topik yang sama (Sarno: 2016). Secara semantik, PLSA menghitung keterkaitan kata dalam dokumen untuk mengidentifikasi adanya kemiripan topik dalam dokumen. Perhitungan PLSA dalam analisis pencocokkan proses bisnis adalah dengan membandingkan label teks pada tiap *node* di bisnis proses. Aspek utama dalam PLSA adalah mendeteksi variabel (*latent*) dalam dokumen yang tidak terlihat atau tidak terdefiniskan (Hofmann:1999).

Efektifitas sebuah model proses bisnis dapat dilihat dari keterkaitan, hubungan antar aktivitas. Menurut Klinkmuller dan Weber (2017), ditinjau dari alur pengendalian informasi [9] dalam model proses dapat menentukan kecocokkan atau kemiripan dalam beberapa model proses yang kompleks. Dengan mendeteksi label-label pada aktivitas di proses bisnis yang mempunyai hubungan yang sama, mengecek konsistensi dalam efektifitas keterkaitan proses, serta membandingkan dengan teknik pencocokkan model lain (Klinkmuller).

Hasil dari penelitian ini adalah mengukur kemiripan model proses dengan perhitungan semantik PLSA dan kemiripan struktural dengan algoritma *Greedy*. Dengan mengukur kemiripan model proses bisnis, maka dapat mengetahui efektifitas dari proses bisnis yang memiliki keterkaitan aktivitas. Teknik pencocokkan graf dengan menghitung semantik tiap-tiap label pada aktivitas yang memiliki keterkaitan dengan label aktivitas lain. Secara struktural, aktivitas-aktivitas yang terhubung memiliki keterkaitan proses atau fungsi yang sama.

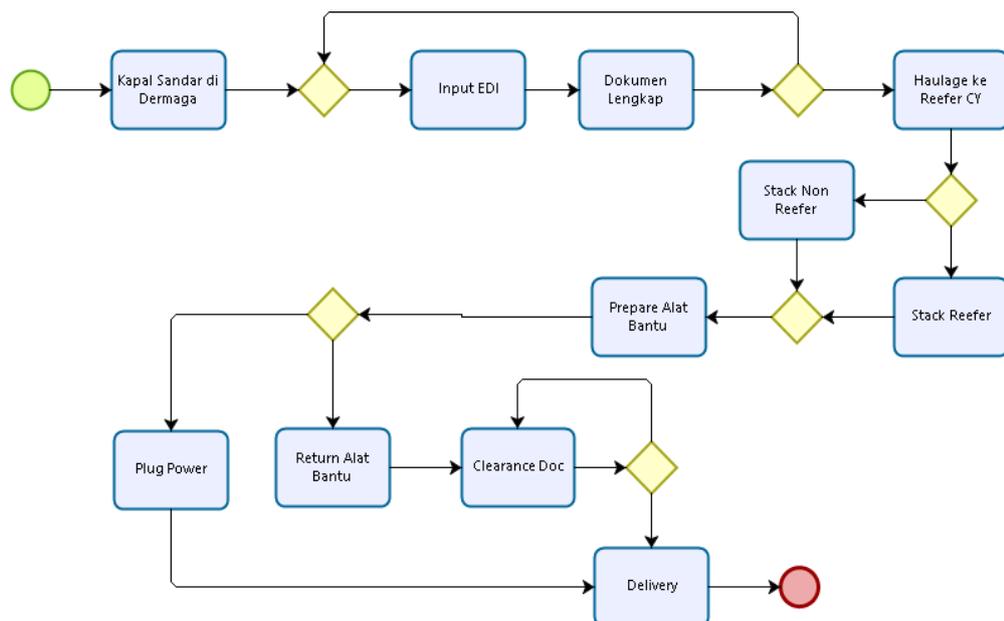
## 2. METODE PENELITIAN

### 2.1. Dataset

Dataset yang digunakan pada penelitian ini adalah kumpulan *event log* dari PT. Terminal Peti Kemas, Surabaya. *Event log* berisi proses pengadaan barang di PT. Terminal Peti Kemas. Setiap prosesnya menjelaskan aktivitas pengadaan.

### 2.2. Model Proses Bisnis

Berdasarkan *event log* yang telah dikumpulkan, selanjutnya adalah mengeneralkan *event log* tersebut ke dalam bentuk model proses. Dalam pemodelannya, penelitian ini mengacu pada model proses dengan notasi BPMN. Pemodelan dengan BPMN pada Gambar 3.3 menggambarkan alur proses dari awal hingga akhir, lengkap dengan aktivitas dan fungsionalitasnya.



Gambar 2.1. Model BPMN Proses Bisnis Pada Peti Kemas

### 2.3. Graph Matching

Teknik (*graph matching*) ialah menghitung kemiripan proses bisnis. Pemodelan graf proses bisnis dari proses pengadaan barang PT. Terminal Peti Kemas Surabaya (TPS) menggambarkan beberapa task atau aktivitas pada tiap *node* nya dan dihubungkan dengan *edge* sebagai alur atau *flow* tiap aktivitas dengan aktivitas lain.

### 2.4. Pengukuran Kemiripan

Pada penelitian ini, pengukuran kemiripan menggunakan dua metode, yaitu berdasarkan perhitungan kemiripan semantic dan struktural dalam mencocokkan graf kemiripan. Kedua metode tersebut adalah perhitungan semantic dengan *Probabilistic Latent Semantic Analysis* (PLSA) dan algoritma *Greedy* untuk struktural *graph matching*.

#### 2.4.1. *Probability Latent Semantic Analysis* (PLSA)

Menghitung kemiripan semantic menggunakan PLSA, yaitu dengan mencari keterkaitan label aktivitas pada model satu dengan model lainnya. Dengan menghitung nilai probabilitas kemiripan label pada tiap aktivitas. Salah satu kelebihan PLSA adalah dapat menemukan kata-kata yang tidak terdefiniskan secara eksplisit. Dalam menghitung tingkat kemiripan semantic antara dua proses bisnis dengan PLSA ialah dengan menggali topik dari masing-masing proses bisnis. Pada proses bisnis, topik digali dari kumpulan *event log*.

Topik diambil dari nilai probabilitas jumlah tertinggi *term* pada kedua dokumen. Dokumen terkait adalah proses model bisnis yang terdiri dari kumpulan *node* atau aktivitas pada proses bisnis. Setelah menentukan topik, kemudian membandingkan topik pada dokumen aktivitas proses bisnis dan label pada SOP (*standar operating procedure*). Pada dataset ini, diambil contoh 3 (tiga) buah topik. Topik-topik ini terkait proses pengadaan pada terminal peti kemas. Hasil perhitungan dari probabilitas *term* pada topik dalam dilihat pada Tabel 2.1.

**Tabel 2.1. Probabilitas *Term* Terhadap Topik**

Term	Topic		
	Topic 1	Topic 2	Topic 3
'kapal'	0,089067	0,036191	0,021144
'sandar'	0,001125	0,099235	0,102357
'dermaga'	0,033369	0,109355	0,049782
'input'	0,007288	0,058656	0,020833
'dokumen'	0,086081	0,039533	0,061976
'lengkap'	0,007583	0,043943	0,069512
'haulage'	0,033477	0,081799	0,039628
'reefer'	0,02326	0,026295	0,109842
'stack'	0,086631	0,007425	0,049952
'prepare'	0,089227	0,014688	0,050411
'alat'	0,08607	0,017552	0,070664
'bantu'	0,111248	0,052988	0,027164
'plug'	0,008168	0,020849	0,000146
'power'	0,024002	0,049656	0,039362
'return'	0,05816	0,022496	0,036793
'clearance'	0,017394	0,10469	0,083742

'delivery'	0,105959	0,037432	0,000292
'berth'	0,031795	0,107035	0,091208
'rantai'	0,100096	0,070182	0,075195

Selain menghitung probabilitas *term* pada topik, probabilitas topik pada dokumen juga dihitung dalam perhitungan PLSA. Tabel 2.2 menunjukkan probabilitas topik pada tiap dokumen.

**Tabel 2.2. Probabilitas Topik Terhadap Dokumen**

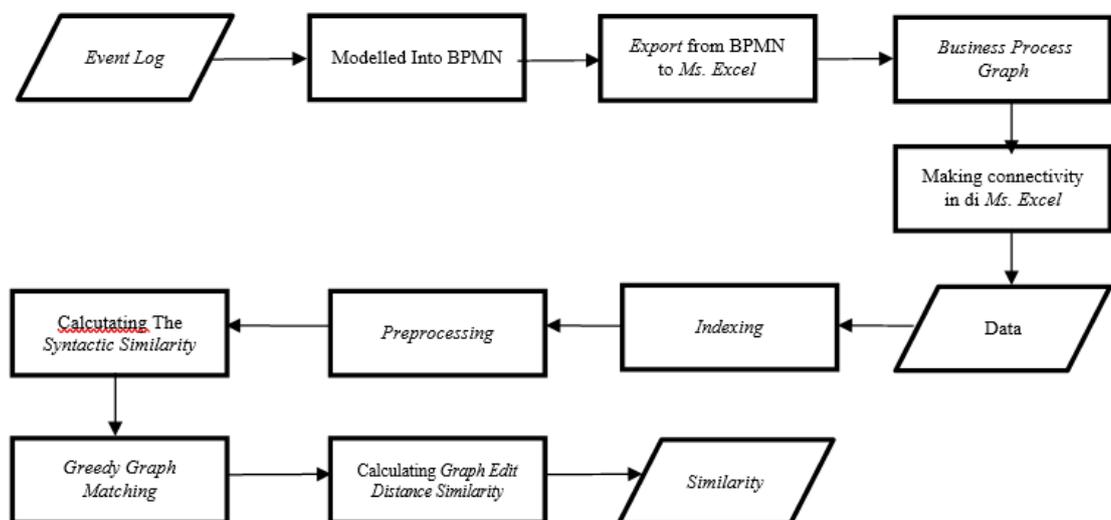
Topic	Document							
	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5	Doc 6	Doc 7	Doc 8
Topic 1	0,104781	1,20E-10	0,017651	0,000132	0,139178	0,125314	0,996152	0,004276
Topic 2	0,756157	0,999987	0,007694	0,234599	2,81E-07	8,25E-08	1,52E-05	0,995286
Topic 3	0,139062	1,26E-05	0,974655	0,76527	0,860822	0,874686	0,003832	0,000439

Dari hasil probabilitas topik, selanjutnya adalah mencari atau menghitung kemiripan antar dokumen dengan menggunakan *cosine similarity*. Sebagaimana rumus perhitungan *cosine similarity* pada rumus (1).

$$CosSim(d_i, q_i) = \frac{q_i \cdot d_i}{|q_i| |d_i|} = \frac{\sum_{j=1}^t (q_{ij} \cdot d_{ij})}{\sqrt{\sum_{j=1}^t (q_{ij}^2) \cdot \sum_{j=1}^t (d_{ij}^2)}} \quad (1)$$

2.4.2. Greedy Graph Matching

*Greedy graph matching* melakukan identifikasi terhadap model-model proses yang mempunyai kemiripan struktur dari model utama proses bisnis. Perhitungan dengan *greedy graph matching* langkah pertama adalah memodelkan sebuah proses bisnis ke dalam bentuk BPMN dari data *event log*. Dilakukan perbandingan model proses utama proses bisnis dengan model *discovery*, pada kasus disini terdapat 4 (empat) model modifikasi. Gambar 2.2. Menggambarkan proses perhitungan kemiripan dengan menggunakan *greedy graph matching*.



**Gambar 2.2. Alur Proses Perhitungan Struktural Dengan Greedy Graph Matching**

### 3. HASIL & PEMBAHASAN

Tujuan dari penelitian adalah dengan menghitung kemiripan model proses bisnis menggabungkan metode PLSA dan *Greedy*, sehingga dapat mengetahui seberapa besar efektifitas dari model proses yang dibuat. Dimana model-model tersebut mempunyai keterkaitan dan hubungan pada masing-masing aktivitas. Maka untuk melakukan analisis dan evaluasi terhadap hasil yang telah dikumpulkan ialah dengan *precision* (P), *recall* (R), dan *f-measure* (F) .

**Tabel 3.1. Hasil Perbandingan Kemiripan Dengan *Greedy Graph Matching***

	V <sub>0</sub>	V <sub>1</sub>	V <sub>2</sub>	V <sub>3</sub>	V <sub>4</sub>
V <sub>0</sub>	1	0.81738	0.89111	0.42902	0.95238
V <sub>1</sub>	0.81738	1	0.71553	0.20792	0.77294
V <sub>2</sub>	0.91828	0.71553	1	0.43627	0.88653
V <sub>3</sub>	0.42902	0.20792	0.43627	1	0.40882
V <sub>4</sub>	0.95238	0.77294	0.85936	0.40882	1

**Tabel 3.2. Analisis Hasil Perhitungan Kemiripan Model Proses Bisnis**

	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
PLSA	82,75%	80%	81,35%
<i>Greedy Graph Matching</i>	76,47%	100%	86,66%

Berdasarkan hasil perhitungan tersebut, model proses dapat dikategorikan efektif bila memiliki tiga (3) kriteria, antara lain :

1. Keterkaitan antar aktivitas dalam satu proses
2. *Flow* atau alur proses yang sederhana
3. Kemiripan atau kesamaan aktivitas

### KESIMPULAN

Efektifitas sebuah model proses bisnis dapat dilihat dari keterkaitan, hubungan antar aktivitas. Model proses dapat menentukan kecocokan atau kemiripan dengan beberapa model proses yang kompleks. Teknik pencocokkan graf dengan menghitung semantik dari setiap label aktivitas yang salinh berhubungan atau keterikatan dengan label aktivitas lain. Pada perhitungan structural, keterhubungan antar aktivitas memiliki proses atau fungsi yang sama. PLSA dan *Greedy Graph Matching* dapat diterapkan dalam perhitungan kemiripan model proses bisnis dinyatakan dengan nilai *precision* dan *recall* diatas 85%. Untuk penelitian selanjutnya dapat dikembangkan lagi dengan menerapkan model *non-free choice* serta tidak hanya mempertimbangkan alur atau *flow* proses sekuensial, namun bisa secara proses parallel dalam menghitung kemiripan model proses bisnis.

### DAFTAR PUSTAKA

- Anugrah, I.G. & Sarno, R. 2016. *Business Process Model Similarity Analysis Using Hybrid PLSA and WDAG Methods*, pp. 231–236.
- Becker, M. & Laue, R. 2011. "Analysing Differences Between Business Process Similarity Measures", in *International Conference on Business Process*.
- Djikman, R., Dumas, M., Dongen, B.V., Reina, K., & Mendling, J. 2011. *Similarity of business process models: Metrics and evaluation*, vol. 36, pp. 498–516.
- Djikman, R., Dumas, M., & Garc, L. 2009. "Graph Matching Algorithms for Business Process Model Similarity Search," in *International Conference on Business Process Management*, pp. 48–63.
- Djikman, R., & Dumas, M. 2009. *Aligning Business Process Models* ", in *Enterprise Distributed Object Computing Conference*, pp. 1–16.

- Hofmann, T. 1999. "Probabilistic Latent Semantic Analysis," *Proc. Fifteenth Conf.*
- Klinkmüller, C. & Weber, I. 2017. *Analyzing control flow information to improve the effectiveness of process model matching techniques*, *Decis. Support Syst.*, vol. 100, pp. 6– 14. .
- Kuss, E. & Stuckenschmidt, H. 2017. "Automatic Classification to Matching Patterns for Process Model Matching Evaluation," in *CEUR Workshop Proceedings 1979*, pp. 306–319.
- Weidlich, M., Djikman, R., & Mendling, J. "The ICoP Framework: Identification of Correspondences between Process Models".