Extracting Common Fragment Based on Behavioral Similarity Using Transition Adjacency Relations For Scalable Business Processes

Abd. Charis Fauzan, Riyanarto Sarno, Muhammad Ainul Yaqin, Amin Jamal
Departement of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indoensia
fauzancharis@gmail.com, riyanarto@if.its.ac.id, yaqinov@gmail.com, amin.jamal@tps.co.id

***Abstract*—Many organizations are already using business process management technology and rapid changes in business processes led to the issue of forming a common business processes becomes very important to be studied. Common fragment is also useful for retrieving business processes that have been stored in the repository. The purpose of retrieving the business process contained in the repository is to compose new business process. In this research, we assume that the business process stored in the repository is a scalable business process using petri net models. This paper proposed method to extract common fragment based on behavioral similarity using transition adjacency relations for scalable business processes. Experimental results show that the proposed method is capable for extracting common fragments based on available business process models. Each petri net model is generated into a common fragment and a variation fragment. Extracted common fragment and variation fragment are useful to compose new business process according to user needs.**

Keywords—common fragment; scalable business process; similarity ; transition adjacency relations ; scalability

I. INTRODUCTION

In recent years, business process management technology is increasingly being used in line with the number of existing business processes [1]. Many organizations are already using business process management technology and rapid changes in business processes led to the issue of forming a common business processes becomes very important to be studied [2]. Common fragment can be defined as a general part of several business processes are interrelated. Common fragment is expected to contribute to creating the business process model that can be adapted to a wide range of possible changes in organizational environment. Common fragment would be a reference to establish a flexible process by which changes can be done easily at any time [2][3]. In addition, the business process model who grew up in the organization also requires effective ability to find a model of the process with good computational approach [4]. Common fragment is also useful for retrieving business processes that have been stored in the repository [5]. The purpose of retrieving the business process contained in the repository is to compose new business process according to user needs [6]. In this research, we assume that the business process stored in the repository is a scalable business process.

Commonly, scalability is an ability of the process or system to handle the growing number of processes or work. For accommodate that growing number, they have the potential process to be enlarged [7]. For business process, scalability be defined as the growth rate among two business process models. A metric of scalability measurements show that the two models have been compared is scalable or not. Measuring the scalability is an interesting research. Previously, Tsai [8] has proposed scalability metric for testing the scalability of software as a service (SaaS) applications. Other hands, the scalability metric that shown in [9][10][11] means that scalability metric has implemented in large areas and an active research topic. This research also proposed scalability metric for models of business process. This paper proposed metric of scalability measure based on similarity metric and complexity measure between two business process models. In the repository, it is possible that it is stored of some scalable business processes. Therefore, extracting common fragment for scalable business process is necessary.

Then, extracted common fragment of business process model is determined by calculating the similarity of some of the business process model, and generated the similar part of all scalable business process. Among the various types of similarity, behavioral similarity is a similarity search efficiently for business process models based on the research of [12][13][14] and using Transition Adjacency Relations algorithm (TAR) for supporting the behavioral similarity. TAR is optimal method to get the value of similarity, which has the advantage of low cost computing [15]. This paper proposes to extract common fragment based on behavioral similarity using transition adjacency relations for scalable business processes.

II. DEFINITION

A. Petri Net Models

Petri net is business process modeling that aims to analyze various models in the business processes [1]. A Model of petri

net has three elements of model, such as transition, place and arc. The transition usually indicates a particular activity (process step or task) that needs to be fire, or a silent step (i.e., t activity) which is used for routing purposes. Place is an element that used to define the pre-conditions transitions and post-conditions transitions. If the precondition is satisfied, a transition can be fired. The result of each firing from transition will be the post-condition. Transitions and places are always connected by directed arcs in such a way that (i) transitions and places have at least one of directed arc and (ii) for each arc, transition can't connected to another transition and then place can't connected to another place [16]. A standard definition of traditional petri nets is presented as below.

A Petri net is a set element (*P, T, F*), where:

- Set of places are denoted by P
- Set of transitions are denoted by T, $P \cup T \neq \emptyset$ and $P \cap T = \emptyset$
- $F \subseteq (P \times T) \cup (T \times P)$ denotes a set of directed arcs that represent a flow relations, joining places and transitions together.

Figure 1 and 2 shows the examples of petri net models of sequence model, parallel model and condition model. In sequence model, there is no gateway included. So, flow of activity in sequence model in figure 1 is (A-B).

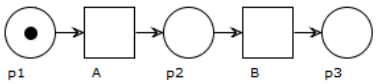

Fig. 1 Examples of Petri Net Sequence Model

Figure 2 (a) is petri net parallel model. Parallel model of petri net includes AND-split gateway and AND-join gateway. So, flow of activity should (A-B-C-D). Based on figure 2 (b), condition model of petri net includes XOR-split gateway and XOR-join gateway. So, the activity in XOR-split gateway has true value if the flow activity has one choice between two choices of activity. Therefore, flow activity has two condition of processes, (A-B-D) or (A-C-D). Traditional petri net has only XOR gateway, AND gateway without OR gateway include.

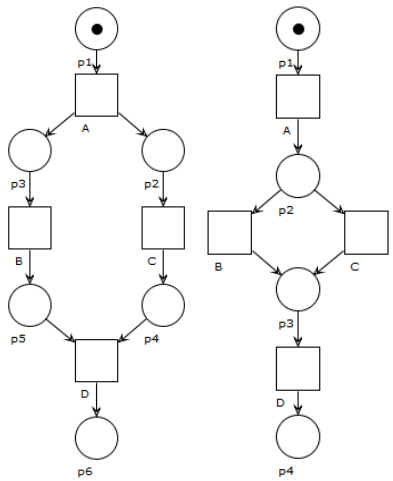

Fig. 2 Examples of Petri Net Parallel (a) and Condition Model (b)

B. Business Process Similarity

Similarity metric among two business process models has aims to measure the similarity value by calculating the similarity according to behavioral or structural similarity method. Some of business processes that have a high similarity value should be arranged in such a way to increase efficiency [17]. Measurement of business process similarity is searching of distance between two business processes were compared. The distance can be calculated by counting the input sequence and the target [19]. Measurement of business process similarity can be done in various ways, that it can use a label matching similarity, structural similarity, or behavioral similarity. Also can be evaluated using three complementary aspects, namely the element labels, graph structure, and execution semantics [20]. Behavioral similarity is a similarity measurement of business processes based on behavioral relationships between business processes [18]. Behavioral similarity using TARs algorithm is obtained from the relation between the existing activities. The relation of this activity can also be interpreted as a sequence of execution that may occur. The term referred as Transition Adjacent Relations (TARs). Behavioral similarity value using TARs algorithm can be calculated by Equation (1). Behavioral similarity using transition adjacency relations is measured by compare the transition adjacency relations (TARs) between two models. The formula of behavioral similarity is mentioned in Equation (1).

$$simB = \frac{(Amount\ of\ Similar\ TARset)^2}{TARset\ model\ 1\ x\ TARset\ model\ 2} \quad (1)$$

For example, in Figure 2, petri has (a) has 3 execution relation, namely AB, BC, CD. Then, for the second net petri has 4 execution relations, namely AB, AC, BD, CD. So, the same number of slices is 2 and each process has 3 and 4 TARset. Then obtained similarity values 4/12 = 0.33.

For Structure similarity, similarity value is measured by compare the similar aspect of petri net models, such as amount of similar place, similar transitions and similar arcs. The formula of structure similarity is shown in Equation (2), for label similarity, it is measured by compare the labels between two models. The formula of label similarity is mentioned in (2) using Jaccard. In this paper, structure and label similarity are used for measuring scalable business process, and behavior similarity using transition adjacency relations is used for extracting the common fragment of scalable business process.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

C. Scalable Business Process

For business process, scalability can be defined as the growth rate among two business process models. A metric of scalability measurements show that the two models have been compared is scalable or not. Metric of scalability is measured

based on similarity metric and complexity measure between two business process models. Business process models that used in this paper are petri net models. Similarity metric is measured by behavioral and structural petri net models, and also using control flow complexity and cyclomatic complexity for measuring complexity of petri net. The Scalability metric determines the connectedness and scalable process between two business processes models. If scalability metric between A and B is 0, determine that A and B are not connectedness models and can't scalable. We proposed a formula for measure the scalability metric between two models of business process based on similarity and complexity measure, mentioned in (8):

$$\psi(A,B) = \frac{\sum C(A)}{\sum C(B)} x\ average(simS(A,B)) \quad (3)$$

Where, measuring S(A) has two alternative formula, by using control flow complexity mentioned in (9) or by using cyclometic complexity mentioned in (10). In section IV, this paper compared both complexities.

$$C(A) = CC(A)\, x \sum Structure(A) \quad (4)$$

- $\psi(A,B)$ is scalability metric between A and B.
- $simS(A,B)$ is structural similarity between A and B.
- $simB(A,B)$ is behavioral similarity between A and B
- $C(A)$ adalah multiplication of complexity by $CFC(A)$ or $CC(A)$ and $Strucutre(A)$, determine the total complexity.
- $CC(A)$ states cyclometic complexity of A, seen from formula (7).

$$CC = |Flow| - |Place| - |Transition| + 2 \quad (5)$$

- $structure(A)$ states amount of petri net element, such as place, arc and transition.

For measuring scalability metric, this paper uses similarity metric to prove that two models have been compared is connectedness or not and using complexity to measure complexity degree of each model.

III. Proposed Method

In this paper, we use 4 example petri net models which representing as scalable business process. Then, using scalability formula in Equation (3), scalability metric will be measured. Scalability metric will get the conditions that minimum value of scalability metric is equal to 0 and also can't less than 0. It occurred because the similarity metric of structure or behavior has a minimal value of at least 0 and maximum value of 1, this result indicates that two compared models has significant different of business process. It also shows that A and B do not have relationship based on their business processes (not connectedness). Finally, it indicates that A and B are not the scalable business processes. Then, miximum value of scalability metric is equal to 1. It shows similarity value also equal to 1. It indicates that A and B has a maximum similarity, and maximum value of scalability metric. And then, the scalable business process will be measured using behavioral similarity based on adjacent transition relations. The steps of behavioral similarity using adjacent transition relations are:

a. Take a transition value from model 1 and model 2
b. Combine two values closest transitional sequentially, on model 1 and model 2. The result is referred as TARset.
c. Calculate the number of TARset model 1 andTARset model 2.
d. Compare TARset model 1 and TARset model 2.
e. Calculate the number of similarTARset between model 1 and model 2.
f. Calculate the similar value using (amount of similar TARset)^2 be devided with (amount of TARset model 1)^2 multiplied with (amount of TARset model 2)^2

Finally, extracted common fragment of scalable business process model is determined by calculating the similarity of some of the business process model, and generated the similar part of all scalable business process. Non common fragment of scalable business process models will be separated as variations fragment.

IV. Discussion

In this section, four models of petri net are shown for testing the scalability metric. There are petri net of model A, model B, model C and model D. We ensure that the existing model is scalable before the common fragment is extracted.

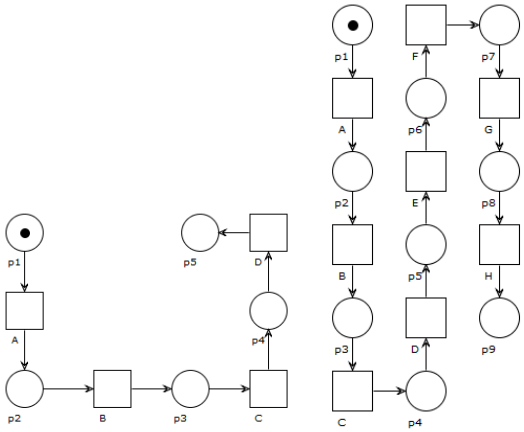

Fig. 3 Petri net model A and petri net model B

Model A and Model B are sequential models, but model A has a smaller business process than model B. Other hand, model C is parallel model with AND gateway. Model D is combination between parallel and condition models with XOR gateway and AND gateway. Scalability metric is determined by measure each model with proposed formula in Equation 3. All of petri net models for extracting common fragment are shown by fig. 3 and fig. 4.

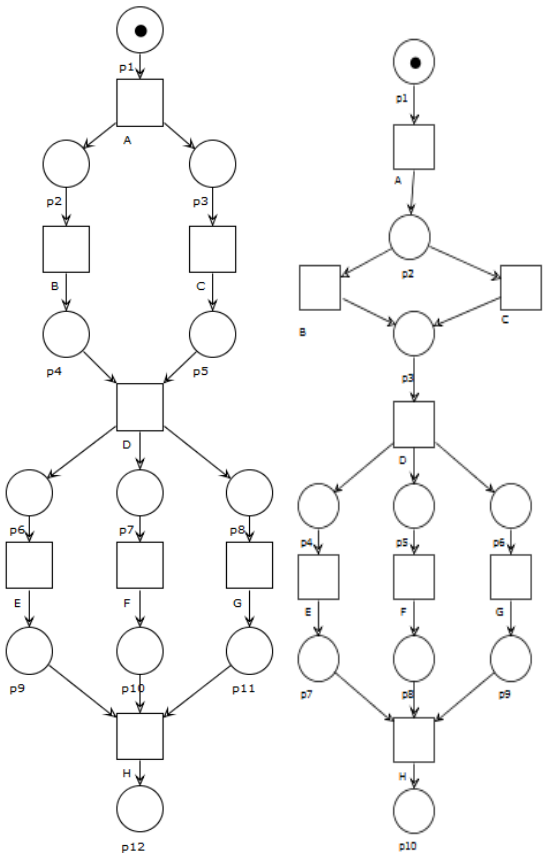

Fig. 4 Petri net model C and Petri net model D

Then, petri net models are measured their scalability metric by calculate the structural similarity and complexity using Equation 2 and Equation 5. This structural similarity determines metric of similarity by structure of each model. The structure of the Petri net model is divided into 3 elements, namely place, transition and arc. The three elements in each netri net model are calculated then compared with other netri net models using jaccard Value of structural similarity among two models are set at intervals 0 to 1. If the similarity value is equal to 0, it means that two models have been compared has significantly different and not connectedness. Vice versa, if the value is equal to 1, it indicates that two models have maximum similar. Metric of structural similarity using Jaccard is shown in table I.

Table I. Metric of Jaccard Structural Similarity

	Model A	Model B	Model C	Model D
Model A	1	0.51	0.31	0.12
Model B	0.51	1	0.41	0.25
Model C	0.31	0.41	1	0.6
Model D	0.12	0.25	0.6	1

Based on the result of similarity metric using jaccard structural similarity (table I), all of models have connectedness for one another. It is obvious that the similarity of structures in table I, has no value equal to 0. It is determined by value of similarity metric, all of them have the value equal to more than 0.

Then, determine of total complexity for each models using cyclometic complexity in Equation 5. For total complexity, it is measured by multiplying cyclometic complexity and amount of petri net element structures. Metric of total complexity C using cyclometic complexity (CC) are shown in Table II, for the scalability metric shown in Table III.

Table II. Metric of C

Model	CC	$\sum structure$	C(model)
C(A)	1	17	17
C(B)	1	33	33
C(C)	4	42	168
C(D)	4	38	152

Table III. Metric of Scalability

	Model A	Model B	Model C	Model D
Model A	1	0.242121	0.022262	0.01398
Model B		1	0.062857	0.052105
Model C			1	0.884211
Model D				1

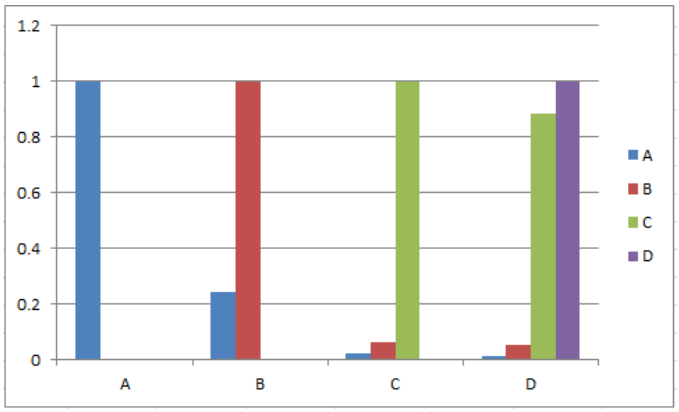

Fig. 5 chart of scalability metrics

Based on metric C using cyclometic complexity (Table II), The calculation is continued to measure metric of scalability using Equation 3. Table III shows the value of metric scalability. Models with low complexity (Model A and Model B) should be compared by models with high complexity to find growth rate between two business process models. Therefore, $\psi(A, C)$ is able to calculate the scalability metric because complexity of A <= complexity of C (see table II). But, $\psi(C, A)$ is disable to calculate because complexity of C > complexity of A. Table III show that all of petri net models are scalable for each other. It is determined by result of metric scalability, there are not value that equal to 0. Highest scalability metric for different models found on $\psi(C, D)$, then lowest scalability metric for different also found on $\psi(A, D)$ Because $\psi(C, D)$ are scalable models, model C can grow into a model D. Otherwise, the two models can't grow if not scalable models. This experiments show that all of petri net model in repository are scalable business processes. The scalable business processes will be extracted for generating common fragment.

Next step, common fragment from scalable business process is extracted based on behavioral similarity using TARs algorithm. Behavioral similarity using TARs determines metric of similarity based on the intended behavior of process models. Automatically, this experiment has been done using java programming. Each process on the Petri net model is divided into two processes. This is known as TARset generation. Based on models of petri net, the value of TARset and amount of tarset are taken for behavioral similarity. Process of behavioral similarity is measured from amount of TARset model 1, amount of TARset model 2 and amount of similar TARset from model 1 and model 2. Table IV shows generated of TARset from each petri net model.

Table IV. Generating TARset of Behavioral Similarity

	Generating TARset	Amount
Model A	AB BC CD	3
Model B	AB BC CD DE EF FG GH	7
Model C	AB BC CD DE EF FG GH	7
Model D	AB AC CD BD DE EF FG GH	8

Based on generated TARset (Table IV), metric of behavioral similarity using TARs algorithm has been measured. Each model of petri net has been compared with the others using Equation 1. Table V shows metric of TARs behavioral similarity.

Table V. Metric of TARs Behavioral Similarity

	Model A	**Model B**	**Model C**	**Model D**
Model A	1	0.43	0.13	0.13
Model B	0.43	1	1	0.23
Model C	0.13	1	1	0.64
Model D	0.13	0.23	0.64	1

For extracting the common fragment, we have to determine the similar TARset of Model A until Model D. In Table VI, the red color on the TARset shows the similar fragment of all petri net model. This is referred to as a common fragment on a scalable business process.

Table VI. Metric of TARs Behavioral Similarity

	Common Fragment
Model A	**AB** BC **CD**
Model B	**AB** BC **CD** DE EF FG GH
Model C	**AB** BC **CD** DE EF FG GH
Model D	**AB** AC **CD** BD DE EF FG GH
Extracted	**AB CD**

After extracted common fragment, we have obtained two TARset to become common fragment of scalable business process. There are TARset of **AB** and **CD**. TARset of **AB** is a combination of processes from A and B. This process will form a new petri net model as the first common fragment (Figure 6). Then, TARset of CD is combination of C and D. This process will form a new petri net model as the second common fragment (Figure 7). Figure 6 and Figure 7 are automatically generated as extracted common fragment for scalable business process.

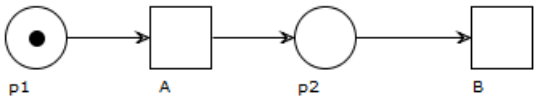

Fig. 6 Extracted Common Fragment 1

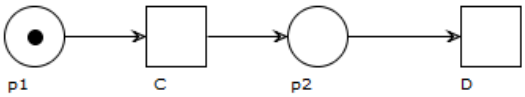

Fig. 7 Extracted Common Fragment 2

Table VII. Metric of TARs Behavioral Similarity

	Common Fragment	Variations
Model A	AB CD	BC
Model B	AB CD	BC DE EF FG GH
Model C	AB CD	BC DE EF FG GH
Model D	AB CD	AC BD DE EF FG GH

Based on petri net models which represented of scalable business process, the common fragment has been extracted. Overall, TARset on the model is divided into two parts. There is a TARset that becomes common fragment. And there are models that become variations. Extracted common fragment of scalable business process which contained in the repository is useful to compose new business process according to user needs. For example, if at any time the user needs model B, then the common fragment will be taken, then variation for model B will be added, and combine with common fragment. .For the next research, we will implement this proposed method for large and variative scalable business process in repository.

CONCLUTION AND FUTURE WORK

This paper proposes a method to extract common fragment based on behavioral similarity using transition adjacency relations for scalable business processes. This paper uses petri net models which representing the scalable business process in repository. Experimental results show that the proposed method is capable for extracting common fragments based on available business process models. Each petri net model is generated into a common fragment and a variation fragment. Extracted common fragment of business process which contained in the repository is useful to compose new business process according to user needs, then, adjusted to the variations required by the user. For the next research, we will implement this proposed method for large and variative scalable business process in repository.

ACKNOWLEDGMENT

The Authors would like to thank the Department of Informatics, Faculty of Information Technology, Institut Teknologi Sepuluh Nopember Surabaya for supporting this research.

REFERENCES

[1] R. Sarno, B.A. Sanjoyo, I.Mukhlash, H.M. Astuti, "Petri Net Model of ERP Business Process Variations for Small and Medium Enterprises," Journal of Theoretical and Applied Information Technology, 54 (1), pp. 31-38. 2013.

[2] R. Sarno, E. W. Pamungkas, D. Sunaryono, Sarwosri, "Workflow Common Fragments Extraction based on WSDL Similarity and Graph Dependency," International Seminar on Intelligent Technology and Its Applications (ISITIA), Pages: 309 – 314, 2015. DOI: 10.1109/ISITIA.2015.7219997

[3] R. Sarno, E.W. Pamungkas, D. Sunaryono, Sarwosri. "Business Process Composition Based On Meta Models," International Seminar on Intelligent Technology and Its Applications (ISITIA). Pages: 315 – 318, 2015. DOI: 10.1109/ISITIA.2015.7219998

[4] R. Sarno and K. R. Sungkono, "Coupled Hidden Markov Model for Process Mining of Invisible Prime Tasks," International Review on Computers and Software (IRECOS), 11 (6), pp. 539-547, 2017. https://doi.org/10.15866/irecos.v11i6.9555.

[5] T. Jin, Jianmin Wang, Marcello La Rosa,Arthur ter Hofstede, Lijie Wen. "Efficient querying of large process model repositories". Computers in Industry,64 (2013) 41–49.

[6] I. G. Anugrah, R. Sarno, R.N.E. Anggraini, "Decomposition Using Refined Process Structure Tree (RPST) and Control Flow Complexity Metrics," International Conference on Information & Communication Technology and Systems (ICTS), 2015. DOI: https://doi.org/10.1109/icts.2015.7379899

[7] Bondi, André B. "Characteristics of scalability and their impact on performance," Proceedings of the second international workshop on Software and performance - WOSP '00. p. 195. 2000. https://doi.org/10.1145/350391.350432

[8] Tsai, Wei-Tek, et al. "Testing the Scalability of SaaS Applications," IEEE International Conference on Service-Oriented Computing and Application (SOCA) 2011. https://doi.org/10.1109/soca.2011.6166245

[9] Y. Chen and X. Sun, "STAS : A Scalability testing and analysis system. in Cluster Computing," 2006 IEEE International Conference on page 1-10. IEEE. 2007. https://doi.org/10.1109/clustr.2006.311882

[10] Montague, Burt., Sexton, Robin. Saas Scalability and Evaluation Table of Content. CSE 564 Software Design : Howework Assigment #1, School of Computing , Informatics, and Decision System Engineering, Arizona State University. 2011

[11] P. Moura, F. Kon, "Automated Scalability Testing of Software as a Service," AST - 8th International Workshop on Automation of Software Test, May 2013, San Francisco, United States. 2013 https://doi.org/10.1109/iwast.2013.6595784

[12] M. Kunze, M. Weidlich, and M. Weske, "A Behavioral Similarity Metric forBusiness Processes," HassoPlattner Institute at the University of Potsdam, Prof.-Dr.-Helmert-Strasse 2-3, 14482 Potsdam

[13] M. Kunze, M. Weidlich, M. Weske, "Behavioral Similarity - A Proper Metric," HassoPlattner Institute at the University of Potsdam, Prof.-Dr.-Helmert-Strasse 2-3, 14482 Potsdam

[14] M. Weidlich, J. Mendling, and M. Weske, "Efficient consistency measurement based on behavioural profiles of process model," IEEE Transactions on Software Engineering, 2010. To appear.

[15] Zha H., Wang J., Wen L., Wang C., and Sun J, "A workflow net similarity measure based in transition adjacency relation," Journal Computer in Industry, Volume 61 Issue 5, pp. 463-471, June, 2010.

[16] Aalst, W.M.P. van der, "The Application of Petri Nets to Workflow Management. Journal of Circuit," Systems and Computers, Vol. 8, No. 1, p. 21-66, 1998.

[17] B. Harjo, R. Sarno, S. Rochimah, "Web Service Similarity with Standarized Description," International Conference on Science in Information Technology (ICSITech). Pages: 115 – 120, 2015. DOI: 10.1109/ICSITech.2015.7407788

[18] R. Dijkman, M. Dumas, B. Dongen, R. Käärik, J. Mendling, "Similarity of business process models: Metrics and evaluation," Information Systems, Volume 36, Issue 2, April 2011, Pages 498-516

[19] Leonard, Michael, et al, An Introduction to Similarity Analysis Using SAS. SAS Institute Inc., Cary, NC. 2005.

[20] M. Dumas, G. Banuelos, L.Dijkman, "Similarity Search of Business Process Models," IEEE Data Eng. Bull. 32(3) (2009) 23-28