Implementation of Process Mining With Flexible Heuristics Miner Algorithm to Support Information System Audit

[1]Aas Suhendar, [2]Gede Agung Ary Wisudiawan, [3]Anisa Herdiani
[1,2,3] School of Computing, Telkom University
Bandung, Indonesia
[1]aas.suhendar@gmail.com, [2]degunk@telkomuniversity.ac.id, [3]anisaherdiani@telkomuniversity.ac.id

Abstract

COBIT 5 as an IT governance framework provides a Process Assessment Model (PAM) that is part of the information system audit process to assess the capabilities of the IT governance process. In the assessment process COBIT 5 there are two important steps namely data collection and data validation. But both steps are still done by taking samples from factual data randomly to be analyzed. It can not yet represent the whole business process because there may be important data that the auditor does not take because it is not a sample data. As a result, the quality of the results of the assessment process becomes less good because it does not use the overall data of business processes in real time. Already the auditor should be able to use the entire data, as in the present era the important data of all business processes are stored in the event log. In this research, process mining with Flexible Heuristics Miner algorithm utilizes event log to get business process model which will be implemented to support the assessment process. The result of research by using event log distribution company obtained a process model that has fitness = 0.983 with threshold parameters DT = 0.6, L1L = 0.98 and L2L = 0.95. After the process model is implemented in the assessment process obtained the result of a rating level point of 63.6% (Large Achieved) and the results of potential bottleneck analysis of the results of enhancement process that can be used for the purposes of audit information system. Thus, the existence of process mining can be applied to support the information system audit process.

Keywords: Information Systems Audit, event log, process assessment, process mining, Flexible Heuristics Miner.

INTRODUCTION

An important activity to ensure that the IT governance process supports business processes is to implement the assessment process [1]. COBIT 5 as one of the IT governance framework has provided Process Assessment Model (PAM) which is part of standardized process information audit process from ISO / IEC 15504-2 to assess the capability of IT governance process [2]. Process assessment model consists of planning phase, data collection, data validation, process attribute rating and reporting. Based on the research [1] there is two important phase of the assessment process that is data collection and data validation. However, both steps of the assessment process are usually still performed by the auditor by collecting sample data from the factual data to analyze the effectiveness of the process, whether it is reliable and in accordance with the expected function [3]. But the randomly drawn sample data can not represent the entire business process because there may be important data that the auditor does not take because it is not a sample data and the data collected are subjective [4].

Consequently, the quality of the assessment process results is poor because it does not use the overall data of business processes in real time [3] [4]. The auditor should be able to use the entire data, as in the present era the critical data of all business processes is stored in event log [4]. Because of these problems then performed a data analysis technique that focuses on the process that is Process Mining. Process Mining is an extraction technique of information from an event log. Process Mining has three types of discovery, conformance, and enhancement [6]. In the discovery of many algorithms that can be applied to get the business process modeling

of the event log, among them is Flexible Heuristics Miner (FHM).

This research uses Flexible Heuristics Miner (FHM) Algorithm in discovery. This algorithm is chosen because it has the advantage of handling event logs that have noise as well as low-structured domains [7]. In this research focus on making process mining as supporting data collection process and data validation in the assessment process and applied to a case study of information system audit in a distribution company [8]. The process assessment model in this research is applied to the domain DSS01 (Manage Operations) on DSS01.01 practices (perform operational procedures), because only in this domain allows for the implementation of process mining [1]. The result of process assessment model in this research is level rating point as result of comparison of SOP process flow with process model from process mining and business flow analysis from event log data which can be used for the purpose of auditing the information system.

THEORY

2.1 Process Mining

Process mining is a discipline that is between machine learning and data mining and process mining is also doing the process of modeling and analysis of a process. The idea of process mining is to find, monitor, and improve the real process by extracting the knowledge of the event logs that are available on the information system. Event logs are data recorded automatically by information systems that describe the series of business processes that are executed within a certain time [6].

Process Mining has three types of discovery, conformance, and enhancement [6] that can be seen in figure 2.1. Discovery is a process for the establishment of a business process model that actually runs based on event logs stored in the information system at the company. Conformance is a process comparison between the process model of discovery with event logs to see the suitability of the model construction. Enhancement is a process for developing and recommending a process model of an existing process model and has been adapted to process modeling of event log data. The usual techniques in enhancement are repair and extension. Repair is to modify the model to better reflect the reality while the extension is to add a new perspective based on the information in the event log [6].

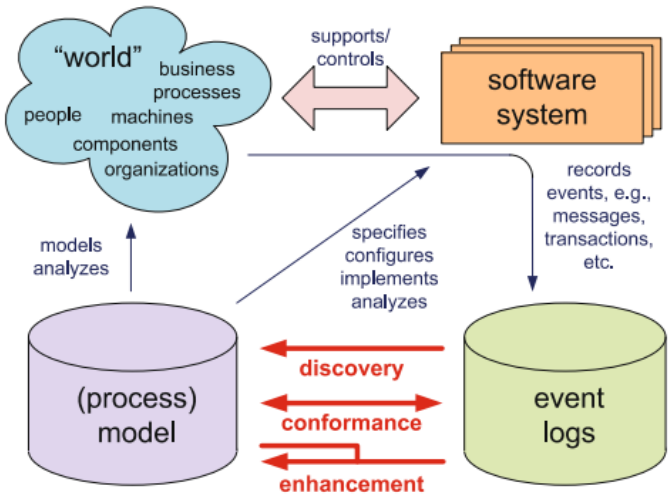

Figure 2.1 Process Mining Overview

2.2 Flexible Heuristics Miner (FHM)

Flexible Heuristics Miner (FHM) in building a process model has an approach by identifying the frequency of interconnection between process and process sequencing in building a model process. There are three threshold parameters available in the FHM to indicate that a dependency relation is received that is Dependency Threshold (DT), Length-One Loop Threshold (L1L) and Length-Two Loop Threshold (L2L). To construct the process model of the event log, the event log should be analyzed for the dependency between events. The process performed to analyze the dependency of events in FHM is to build a Dependency Graph. Dependency Graph is a model that represents the dependency between events on the relation there is a value of dependency that is calculated by the dependency measure. Dependency measure consists of dependency event (2.1), Length-One Loop dependency (2.2), and Length-Two Loop dependency (2.3) [7].

$$a \Rightarrow_w b = \left(\frac{|a >_w b| - |b >_w a|}{|a >_w b| + |b >_w a| + 1} \right) \quad (2.1)$$

$$a \Rightarrow_w a = \left(\frac{|a >_w a|}{|a >_w a| + 1} \right) \quad (2.2)$$

$$a \Rightarrow_{2w} b = \left(\frac{|a \gg_w b| + |b \gg_w a|}{|a \gg_w b| + |b \gg_w a| + 1} \right) \quad (2.3)$$

Let *W* be an event log over *T,* and a,b $\in T$. Then $|a >_w b|$ is a number of times $a >_w b$ occurs in *W*. For Length-One Loop and Length-Two Loop let *W* be an event log over *T,* and a,b $\in T$. Then $|a >_w a|$ is a number of times $a >_w a$ occurs in *W*. And $|a \gg_w b|$ is a number of times $a \gg_w b$ occurs in *W*.

There are many techniques performed to perform conformance, one of which is to calculate the fitness of a process model. A process model has good fitness quality if all traces in the event log can be parsed with the process model [9]. In literature [5] there is a way of calculating fitness with the method of Continuous Parsing Measure (CPM) (2.4), where *e* is a number of events, *m* is a number of missing event and *r* is a number of remaining event.

$$CPM = \frac{1}{2}\frac{(e-m)}{e} \times \frac{1}{2}\frac{(e-r)}{e} \quad (2.4)$$

Method

An overview of the system can be seen in Figure 3.1 which refers to the COBIT 5 model assessment process.

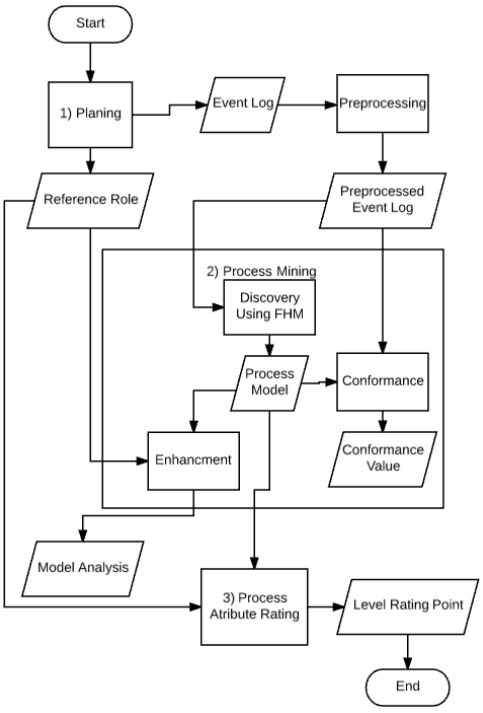

Figure 3.1 General Flow System

3.1 Planning

The planning phase describes all activities to be undertaken in the assessment, identifies the scope of the process, determines the information documentation like standard oprational procedure document and event log data. In this reaseach applied to the domain DSS01 (Manage Operations) on DSS01.01 practices (perform operational procedures).

3.2 Preprocessing

The event log will be preprocessing to simplify and equate the format to fit the system design. Preprocessing is done by clearing some unused columns. The data required by the system is the id of the case, the name of the activity, and the time and user executor of the activity.

3.2 Process Mining

For support data collection process and data validation in the assessment process implemented process mining in this phase. On discovery process in this research using Flexible Heuristics Miner (FHM) algorithm. The Event log that has been done preprocesing then formed into a process model. The resulting process model that will be used in the conformance process to see how to fit the model generated in parse the event log. Then the process model will also be used in the enhancement process for bottleneck analysis. The results of process mining are the discovery process model and bottleneck analysis from event log data. Discovery process model is constructed with Flexible Heuristics Miner (FHM) algorithm that has three parameters, they are Dependency Threshold (DT), Length-One Loop Threshold (L1L) and Length-Two Loop Threshold (L2L).

The parameters used will have an impact on the resulting model. The challenge in process mining is that we must optimally use these parameters to avoid overfitting process model (process model is too detailed) and underfitting process model (process model are too general)

3.3 Process Attribute Rating

Process attribute rating is the assessment phase and the attribute of the rating point level to all processes in the domain being audited. This rating is based on validated data at the data validation phase in the assessment process. In this reaseach, process attribute rating used to the domain DSS01 (Manage Operations) on DSS01.01 practices (perform operational procedures). The process attribute rating is done by performing a process of comparison between the process model produced by FHM with SOP process model. This comparison is done using Set Theory, which is the ratio of a number of intersection and number of union between the two process modeles. The similarity of two process models gets the capability level that can be seen in Table 3.1.

Table 3.1. Capability Level of Process Attribute Rating

Level	Achievement
N : Not Achieved	0 % - 15 %
P : Partially Achieved	15 % - 50 %
L : Largely Achieved	50 % - 85 %
F : Fully Achieved	85 % - 100 %

Result & Analysis

Based on the results of tests conducted by entering different threshold (Dependency Threshold (DT), Length-One Loop Threshold (L1L) and Length-Two Loop Threshold (L2L)) values to determine the most appropriate threshold value in generating a process model with a good fitness value is DT = 0.60, L1L = 0.98 and L2L = 0.95. The threshold value is taken because with the value of the resulting model has fitness 0.983 which means the model is almost able to parse most of the trace in the event log and represent the model in accordance with the events in the event log. With the threshold values are also obtained a process model that is not too overfitting (model process is too obvious) and underfitting (process model is too general).

4.1 Result of Process Attribute Rating

After obtaining the optimal threshold value that produces the fitness value 0.983, the resulting process model is then performed a comparison of the process flow of the SOP described in Figure 1

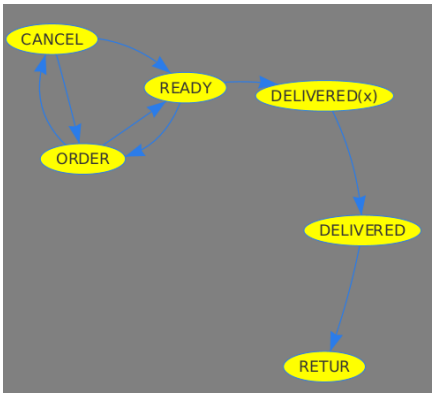

Figure 4.1 SOP Process Model

The result of process model generated by process mining can be seen in Figure 4.2 When compared with the SOP process model then get the rating point value level of 63.6% to obtain L (Largely Achieved) capability level.

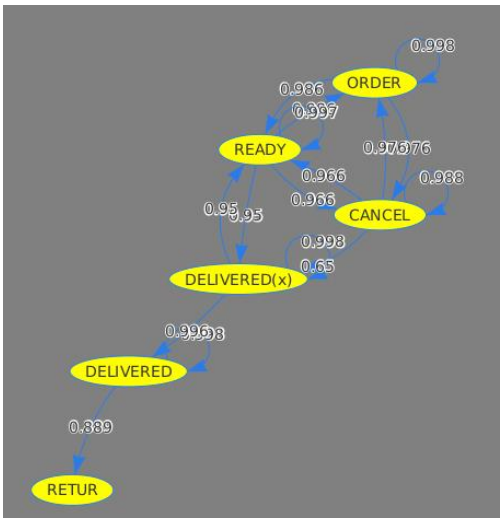

Figure 4.2 The Result Process Model 1

From the result of the attribute rating test, process model and SOP almost have similarities although there are some event dependencies that are not reflected in the SOP when in the event log is executed like dependency event ORDER => ORDER, READY => READY, CANCEL => CANCEL, DELIVERED (x) => DELIVERED (x) , READY => CANCEL, DELIVERED (x) => READY, CANCEL => DELIVERED (x). Because there are findings of process models that are not in accordance with the SOP to eat will affect the value level point value obtained.

4.2 Analysis of Bottleneck Process

The bottleneck analysis was performed on the average time of the longest activity that it was suspected to cause a bottleneck. Here is the time table in Table 4.3 for the average in the main flow process in the company.

Table 4.3. Bottleneck between activity

Start Event	Destination Event	Frequency	Mean Time (Minute)	Mean Time (hour)
ORDER	READY	315	420.39	7,0065
ORDER	CANCEL	86	326.18	5,43
READY	DELIVERED(x)	245	309.3	5,15
DELIVERED(x)	DELIVERED	240	565.99	9,408
DELIVERED	RETUR	8	4240,87	70,68

The dependency of events in flow according to SOP made from the side of the company contained in Table 4-16 consist of ORDER => READY, ORDER => CANCEL will cause the potential bottleneck because the frequency of occurrence is high and the time required by the dependency event is long enough. Likewise READY => DELIVERED (x), DELIVERED (x) => DELIVERED will cause potential bottleneck even though it is likely that many other factors are inhibiting the auditor's need to be discussed with the company. With the analysis of potential findings, this bottleneck will assist auditors in the process of auditing the information system.

Conclusion

Based on testing done in this research, it can be concluded that process mining with Flexible Heuristics Miner can support the data collection and data validation phase in the process assessment model. And in this research is implemented process mining at event log distribution company with parameters used in FHM in the form of Dependency Threshold (DT) = 0.6, Length-One Loop Threshold (L1L) = 0.98 and Length-Two Loop Threshold (L2L)

= 0.95 can generate process models that represent business processes in event logs with Fitness 0.983 of total trace = 337 and event count = 3461.

Process attribute rating on COBIT 5 DSS01 (Manage Operations) domain in DSS01.01 management practice (perform operational procedures) with comparing between process model and SOP from distribution company obtained rating point value of 63.6% with level L (Large Achieved) with the use of threshold that has been tested. And then enhancement in process mining with Flexible Heuristics Miner (FHM) can detect business process bottleneck in the form of an inter-process bottleneck which is useful to develop the running business process. These results are useful for auditors in the process of auditing the information system.

REFERENCES

[1] A. P. Kurniati and I. Atastina, "Implementing process mining to improve COBIT 5 assessment program or managing operations (Case study: A university blog)," ResearchGate, vol. 72, no. 2, pp. 191–198, Jan. 2015.

[2] ISACA, Process Assessment Model (PAM) : Using COBIT 5. ISACA, 2012.

[3] W. M. P. v Aalst, K. M. v Hee, J. M. v Werf, and M. Verdonk, "Auditing 2.0: Using Process Mining to Support Tomorrow's Auditor," Computer, vol. 43, no. 3, pp. 90–93, Mar. 2010.

[4] A. P. Kurniati, G. P. Kusuma, and G. A. A. Wisudiawan, "Designing application to support process audit using process mining," ResearchGate, vol. 80, no. 3, pp. 473–480, Oct. 2015

[5] A. J. M. M. Weijters, W. M. P. Aalst, and A. A. K. Medeiros, "Process Mining with the Heuristics Miner-algorithm," ResearchGate, vol. 166, Jan. 2006.

[6] W. M. P. van der Aalst, Process Mining - Discovery, Conformance and Enhancement. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[7] A. J. M. M. Weijters and J. T. S. Ribeiro, "Flexible Heuristics Miner (FHM)," in 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), 2011, pp. 310–317.

[8] I. Tawakal, "Analisis dan Implementasi Audit Menggunakan Process Mining Algoritma HeuristicMiner dengan DSS01 COBIT 5 (Studi Kasus: CV. Narnia Distribution)." Universitas Telkom, 2016.

[9] W. van der Aalst et al., "Process Mining Manifesto," in Business Process Management Workshops, vol. 99, F. Daniel, K. Barkaoui, and S. Dustdar, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 169–194.