# Applying Process Mining to the Ship Handling Process at Oil Terminal

1st Julia Rudnitckaia
*Faculty of Information Technology*
*IT4Innovations - Center of excellence Brno*
*Brno University of Technology*
Brno, Czech Republic
irudnickaia@fit.vutbr.cz

2nd Wacharawan Intayoad
*IGS - International Graduate School for Dynamics in Logistics*
*University of Bremen*
Bremen, Germany
int@biba.uni-bremen.de

3th Till Becker
*Faculty of Business Studies*
*University of Applied Science Emden/Leer*
Emden, Germany
till.becker@hs-emden-leer.de

4rd Tomáš Hruška
*Faculty of Information Technology*
*IT4Innovations - Center of excellence Brno*
*Brno University of Technology*
Brno, Czech Republic
hruska@fit.vutbr.cz

***Abstract*— Logistics processes are characterized by a high degree of complexity, dynamics, diversity, and human-centricity. The reasons are manifold - product complexity, unexpected events (such as strikes, weather conditions, accidents, etc.), and the high number of participants in the logistics processes. Improving our understanding and making the process more transparent can help to avoid risks and increase performance. The main purpose of this paper is to show a practical application of process mining in the specific domain of seaport terminals. We apply process mining techniques to real port data and discover process models without a priori knowledge. A new framework for the improvement of process mining results using context-aware information in terms of seaports is proposed.**

***Keywords— logistics processes, seaport terminal, process modelling, context awareness***

## INTRODUCTION

Logistics is a constantly developing field. It is more and more insufficient to improve only infrastructure of transshipment points and to make decisions based on established statistics methods as the amount of transshipped goods has grown significantly in recent years. Nowadays, there are prerequisites for using more powerful mining methods. Port Communication systems, the single window concept, fully automated terminals (e.g. Maasvlakte2 in the Netherlands), embedded ERP and other information systems can provide stakeholders with necessary information about ongoing processes. Simultaneously, they provide various kinds of collected data which can be used to acquire valuable knowledge that can be used further for process monitoring, optimization, bottlenecks detection, forecasting, and risk assessment.

The process mining approach offers new and more effective options to deal with the above-mentioned challenges. They extract knowledge from event logs collected during activities in business processes and present the discovered knowledge in the form of process models. A business process is composed of various activities or tasks that creates values for products and services. Dynamic change in the business environment requires adaptive and flexible business processes in order to compete in the market. As a result, the actual business processes are not strictly following a single process model. Business processes in the real world have a very high number of variants in order to achieve customer needs.

The discovered process model from process mining reflects a real process model (descriptive model) and its variants with the possibility of extension (scaling process), determination bottlenecks (traffic jam), detection of deviations to enable operational responses, and representation of different perspectives (control-flow, resources, performance). Also, the model can be used as a basis for prediction and decision-making systems. Advantages of real process models in terms of seaport logistics are as follows:

- Process models make the process transparent for customers, stakeholders, port authority, and other participants of the logistics process. Process transparency can improve our understanding of the chain of the nautical handling process of maritime ships around seaports.
- Performance analysis does not only detect bottlenecks within the process but also determines the causes of their occurrence. Therefore, analysis of the discovered process model can lead to a decrease in waiting time and the intensity of the exploitation of available port resources, which also affects the processing cost.

Process mining has been deployed successfully in many studies and real application in various domains such as healthcare [1], software development [2], and education [3]. However, there is a limited amount of studies regarding process mining in logistics and only a few cases in logistics application. Therefore, this research objective is to discuss the application of process mining, its advantages, as well as the challenge in the logistics domain. First, we present the process mining state of the art in Section 2. In Section 3, we discuss the application of process mining in logistics processes from exploring the related works. Afterwards, we evaluate the application of process mining technique using a real case from logistics in Section 4. In Section 5, we draw a conclusion and present a framework for the future research direction of process mining in the logistics.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE
978-1-5386-8500-6/19/$31.00 ©2019 IEEE

## PROCESS MINING

Process mining is a multi-disciplinary approach which encompasses several domains such as computer science, business process management, and mathematics. The goal of process mining is to extract knowledge from event logs to discover, monitor, and improve real processes. In business practice, prescribed models often deviate from what actually happens [4].

Process mining is categorized into three main types: (1) process discovery, (2) conformance checking, and (3) enhancement.

Process discovery is the fundamental task for process mining. Process discovery techniques generate process models reflecting the actual processes. These models provide specific information such as the process bottleneck, the delays, or process deviation [5]. In theory, process discovery is highly useful for flexible environments as it provides valuable information about the actual process at hand. There are several process discovery techniques such as alpha-mining, heuristic mining, fizzy mining, and inductive mining. However, many studies reviewed that process discovery methods encounter challenges when they have to deal with complex and flexible business processes [6], [7].

While process discovery aims at automatic extraction of a process model from event logs, process conformance focuses on the alignment between the actual process and its model. Several studies reveal that models often deviate from reality. For example, in order to react in urgent situations, processes require flexibility to take corrective actions [4]. Conformance checking illustrates where the actions and activities are altered. This kind of information can be used to improve business processes to be more corresponding with the real world. Meanwhile, in some cases, is it crucial to assure that the actual procedure is aligned with the model such as detecting abnormal processes or fraud. In this context, conformance checking methods could reveal alignment between actual procedures and models.

The third type is enhancement. The discovered models can be used in further analysis such as finding bottlenecks and optimizing resource allocation. Furthermore, combining process mining methods with other methods is possible to achieve the objective of the task. The concept of data mining, statistical, and machine learning can be integrated to discover specific questions. For example, standard classification methods can be applied to find the decision point in a process model [8]. Moreover, integration with other methods can be used for prediction and recommendation at runtime.

### A. *Event log*

Even logs are the input for process mining methods. Today's businesses produce an enormous amount of data, particularly generated by the use of information systems. These information systems automatically record activities during operational works in event logs. Event logs are sometimes also called historical data, audit trail, transaction log, etc. An event log contains sequences of events and their properties.

Each event refers to an activity which is related to a single process, often called as 'case' or 'process instance'. An event generally includes time stamps of the event such as start time and end time. The event within a case needs to be ordered, normally by time stamp. It may include additional information such as originator, resource, cost, etc. Assuming that an event log represents a particular process, we can describe an event log as follows;

- A process consists of cases, and each case can be identified by 'case ID'
- A case consists of at least one event, and each event only relates to one particular case.
- Each event refers to an activity.
- Events in a case must be ordered.
- Process properties refers to additional information which are related to process instance.

In this study, we use context-aware information which refers to process properties and other additional information related to the behavior of the business process.

### B. *Process mining challenges*

There are some issues regarding process discovery algorithm. The first one is noise and incomplete data as a common issue. In this context, noise data does not mean incorrect data. Noise refers to the infrequent behavior or outliers. Process discovery techniques should be able to present the majority of traces and filter noise. The approaches which are robust to noise are heuristic mining, genetic mining, and fuzzy mining [9]. On the contrary, incomplete data refers to having too little data. Similarly, in data mining and machine learning, we cannot assume to have seen all possibilities from training data, particularly when amount of training data is limited.

Process models can be evaluated by various dimension. Commonly, there are four dimensions for the determination of the quality of the resulting models: fitness, precision, generalization, and simplicity [9] . In this study, we do not consider conformance checking and the obtained process model was evaluated based on domain expert knowledge.

### C. *Fuzzy mining*

There are some popular algorithms for discovering process models such as alpha-algorithm [9], heuristic mining, genetic algorithm [10], fuzzy miner, etc.

Fuzzy mining is related to heuristic mining. It aims to extract less-structured and complex processes. The existing approached of process mining like alpha-mining and heuristic mining encounter problem when they have to deal with processes from a restrictive environment as it is often found in reality.

In this study, fuzzy mining is deployed for our use case as it is robust to less-structured and complex processes. As a result, fuzzy mining is capable of providing understandable process models from dynamics logistics processes. In contrast, alpha-mining and heuristic mining often encounter problems when they have to deal with processes from flexible environments. The resulted models from these algorithms when mining these types of process are unstructured and hard to comprehend, resulting in so-called "spaghetti models". Fuzzy mining provides a high-level view on process and abstract from undesired details.

Fuzzy process mining involves in identifying two fundamental metrics, significance and correlation [11]. The significance metric takes the frequency of events into account.

The events which are more frequent are more important. The correlation metric is used to determine how closely related two events following one another are. The highly significant activities are contained in the simplified model. The less significant activities but highly correlated are integrated and present as a cluster in the simplified model. The less significant and weakly correlated activities are removed from the simplified model.

Three metrics are used for measuring significance and correlation: unary significance, binary significance, and binary correlation. Unary significance is composed of two primary metrics.

The first one is frequency significance which measures how often a certain event class was observed in event log. Another metric is routing significance. It helps to identify important routing nodes. For instance, the higher the number of the difference between the incoming arcs and outgoing arcs, the more important the node is in terms of routing.

Binary significance relates to a precedence relation between two event classes, which can be used for selecting the edges that will be included in simplified process models. The frequency metric can be used to identify the relationship between two event classes. The more frequent two classes are observed in event logs in a sequence, the more significant their relationship is. Another metric that can be used to identify relationship is distance significant that is used for isolating behavior of interest.

Binary correlation measures the distance of events in a precedence relation, for example, how closely related two events following one another are. Binary correlation is the key indicator used for the decision between aggregation and abstraction of less-significant behavior.

Let $N$ be the number of nodes in a process model, and let the matrix $sig\colon N \times N$, be a relation assigned to each pair of nodes $A, B \in N$. The relative significant of their ordering relation is as follow.

$$rel(A,B) = \frac{1}{2}\left(\frac{sig(A.B)}{\sum_{X \in N} sig(A,X)}\right) + \frac{1}{2}\left(\frac{sig(A.B)}{\sum_{X \in N} sig(X,B)}\right) \quad (1)$$

Relative significance can be used to determine the offset between both (A,B) and (B,A) relations' relative significances. For example, the offset gap can be express as follow.

$$ofs(A,B) = |\, rel(A,B) - rel\ (B,A)| \quad (2)$$

Relative significance, offset between $A, B \in N$, and thresholds can be used to identify the relationship between nodes and edges of the process models.

For example, if $rel(A,B)$ and $rel(B,A)$ exceed a specified preserve threshold value, (A,B) and (B,A) will be preserved for the simplified process models.

Or, in the case of one of conflicting relation's relative significance is below this threshold and $ofs(A,B)$ also exceeds the specific ratio, the relative significance of both conflicting relations differs. Thus, we can remove the less significant relation out of the simplified process model.

Otherwise, i.e. if at least one of the relationships has a relative significance below the preservation threshold and their offset is less than the ratio threshold, both relations can be determined as concurrent. However, both edges are removed from the simplified process models as they do not meet the factual ordering relation. The algorithm is implemented in various software such as Disco, Prom, Celonis.

## RELATED WORKS

Logistics industries require processes analysis for continuous improvement due to the increase of the flexibility requirements from the end customers. There are several cases where process mining is deployed in order to take a decision.

In the work of [12], the authors proposed the methodology for process mining to acquire logistics intelligence for a Chinese bulk port. They gathered event logs from logistics system as well as additional information to enhance the knowledge in other perspectives. As logistics data may not always come readily formatted from workflow management systems, initial steps before applying process mining techniques are required, such as log extraction and preprocessing data. Afterward, process mining techniques were deployed for process discovery, conformance checking, and process performance analysis. The results presented that process mining is capable of providing opportunities for decision support for logisticians. Process mining can be used to enhance logistics process transparency, support the internal control of logistics firms, and improve logistics performance. As well as in the work of [13], process mining algorithms were deployed to discover the actual process of ship entry processing at the oil terminal. The discovered process models revealed the deviations and the bottlenecks which are used for process improvement and for finding solutions. While most of the studies were devoted to enhancing process transparency, the work of [14] explored the relationship between business processes and process properties from the event logs. The authors deployed Naive Bayesian to analysis the relationship between process properties (e.g., the number of process instances which compete for the same resources and the process lead time of the previous process instance) and process performance (process lead time). They concluded that process properties could be useful for enhancing process knowledge. However, in order to answer specific questions, selecting the adequate process properties is crucial.

## USE CASE

### *D. Scenario*

The matter of process optimization in logistics through new mining methods remains relevant. The reasons for applying process mining include the following: reducing the idle time of ship handling, improving of key performance indicators, maximizing the employment of port resources, etc. There are usually basic patterns of work and people from industry are not willing to waste their time without previously known results. Analyzed process models strongly depend on different scenarios in logistics. In case of transport logistics, we can define different cases such as ships, warehouses, containers, documents, forwarders, terminals. For our study, the ship handling process was chosen to present possibilities of applying process mining techniques.

The ship handling process at an oil terminal is supposed to be clear and straight. However, it remains a black box for analysts. There are many simulations and normative models to understand what the ship-handling process is, but most of them do not focus on reflecting the real/descriptive model. Using algorithms of process mining and data from timesheets, the real process model can be constructed. The real process model can be used for audit (comparing normative and real models, finding deviations), training new workers, creating ship-scheduling, prediction of threats, detection of bottlenecks and their causes, and also for recommendation of countermeasures.

## E. Event log

An event log usually contains four important parameters: case id, activities, and timestamps (start and end time for the activity). For further analysis, an event log can be enriched with an information about process attributes such as costs, operators, ship sizes, weather conditions, etc. In our case, an event log (Table 1) is created based on the information from port documents – timesheets/statements (i.e. a detailed chronological description of the activities of the vessel during the stay in a port: arriving, taking the sea pilot on-board, hailing in, mooring, preparing for the loading and unloading operations, the actual loading and unloading operations, unmooring, departure). For the ship-handling process, every ship case at the terminal is referred to a case. While the operations are activities and timestamps are start and end times. Process mining techniques simulate process models, extracting necessary information from event logs.

TABLE I. FRAGMENT OF THE EVENT LOG FOR SHIP-HANDLING PROCESS

| Case ID/ Ship case | Events ID | Activity/Operations | Start time | End time |
|---|---|---|---|---|
| 9997 | 1 | Arrival | 02/01/2012 00:00 | 02/01/2012 08:30 |
| 9997 | 2 | Nor tendered | 05/01/2012 00:00 | 05/01/2012 00:01 |
| 9997 | 3 | Piloting for mooring | 06/01/2012 18:35 | 06/01/2012 19:15 |
| 9997 | 4 | Mooring maneuvers | 06/01/2012 19:15 | 06/01/2012 19:50 |
| … | … | … | … | … |
| 9997 | n | Pilotage for leaving | 07/01/2012 23:20 | 07/01/2012 23:50 |
| 10000 | n+1 | Arrival | 03/01/2012 00:00 | 03/01/2012 06:30 |

## F. Filtering data.

The event log (see Table 1) is obtained from timesheets documents that comprise required information about full ship handling process at an oil terminal for subsequent financial settlement between the ship and the port.

The data was pre-processed in order to avoid the following problems:

- Removing noise: ship cases with start activity "ship arrival" and end activity "Pilotage for leaving" are defined as a process with a normal behavior. The ship cases that do not belong to a confidence interval are referred to noise and outliers (for example, ship cases with turnaround time less than 12hours or more than 22 days). In this way, problems with incomplete cases and noise in data were partially solved. Despite the data preprocessing, the results should be interpreted with caution, as there can be remaining noise in the data.
- Duplicated labels: Raw data is obtained from an old database and does not have a unified standard. For instance, two similar operations «Weather conditions» and «Weather conditions (Other)» (or «NOR tendering-berth» and «NOR tendering-road») have two different identifications. This problem could be tackled by aggregation of operations according to experts' knowledge or based on international standard. For example, two different operations «Mooring with Operator1» and «Mooring with Operator2» can be combined into one operation «Mooring» and adding additional resource parameter Operator. Reducing the number of operations improves quality and representation of final process model.
- Another issue of equal importance is uncertainty in the data. There are duplicates with different timestamps within a case, which may affect the results of algorithms. For example, we can find two the same operations in one case with start and end times (1:00; 2:00) and (1:00; 3:00) respectively. This uncertainty can be solved by help process owners or using the common time frame that covers both times (for our example, the operation has time (1:00; 3:00)).

## G. Discover the process model

To provide insights into the process structure, a process model is presented below. The event log is composed of 2,659 ship cases, 62,245 events and 135 different operations/activities for four years. We took data for one year to discover the process model. Eventually, there were 324 cases and 6,405 events which means each ship case has on average approximately 20 events. In case we show all operations, the discovered model will be complex and unreadable. Therefore, Figure 1 presents the most frequent activities in the process. (Since the analyzed terminal doesn't have any standards for operation names, only frequent operations were translated and others are labeled with original names). We chose the Fuzzy miner as a suitable algorithm for discovering the logistics process model. The process is presented by help an oriented graph with nodes-activities and arcs-relations between them. The first operation is always Ship arrival, next is Nor-tendered, and the whole process is finished with the operation Pilotage for leaving.

Numbers on arcs indicate the number of different ship cases, which go through this part of the process. Thick arcs, in turn, show the most common way for all cases. Moreover, nodes of the diagram detect frequency of operations with colors: grey nodes are less frequent operations and blue ones are the most frequent. It is important to note that in the obtained process model infrequent activities have been removed and only the main flows are presented.

The process model was evaluated by dispatchers and process owners. Therefore, we did not use Conformance Checking in this study.

# RESULT FRAMEWORK

When building complex and unstructured processes like logistics processes the analysts can obtain "spaghetti models" in the end. The incomprehensible and unreadable model cannot be used for representing real process flow to stakeholders or for enhancement with additional process attributes. For this specific issue of process mining, we offer a framework for logistics processes (see Fig.2) which can help to avoid complexity in process models. We offer to use context information according to two main types of process mining – discovery and enhancement. By "context-aware information" we mean characteristics which describe the process instance. It can refer to a person, place, time, frequency of events, operators, and other attributes of the process. In this article, we follow a direct way in the framework, which is marked with a solid line. Two dotted lines show possible solutions for improving process mining techniques in terms of logistics processes.

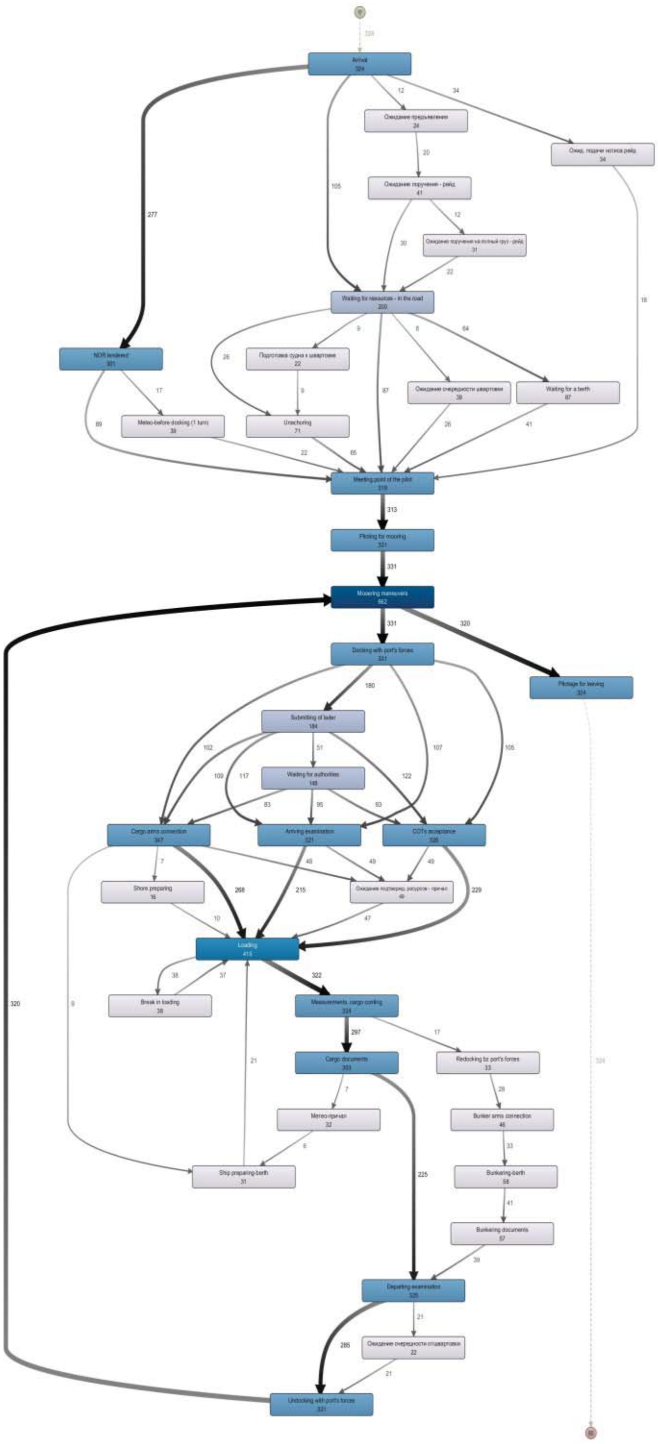

Fig. 1. A Common process model of the ship-handling process at the oil terminal with the most frequent activities (Fuzzy Miner algorithm)

Discovery type: Process Mining techniques are flexible but strongly depend on input data. Therefore, context-information can be used to detect sub-processes, find relationship among process instances, and enhance the result of discovered models in various perspectives. For example, in our case, we propose to use such additional parameters as lay time (i.e. the amount of time specified in a charter party for a vessel's loading and unloading) for building cargo handling process or information about documents type to discover document-flow process. Then, all the sub-process models can be used for making different perspectives of the general process model and avoiding “spaghetti-like models”. Moreover, context information can provide specific point of view for a certain objective of analysis.

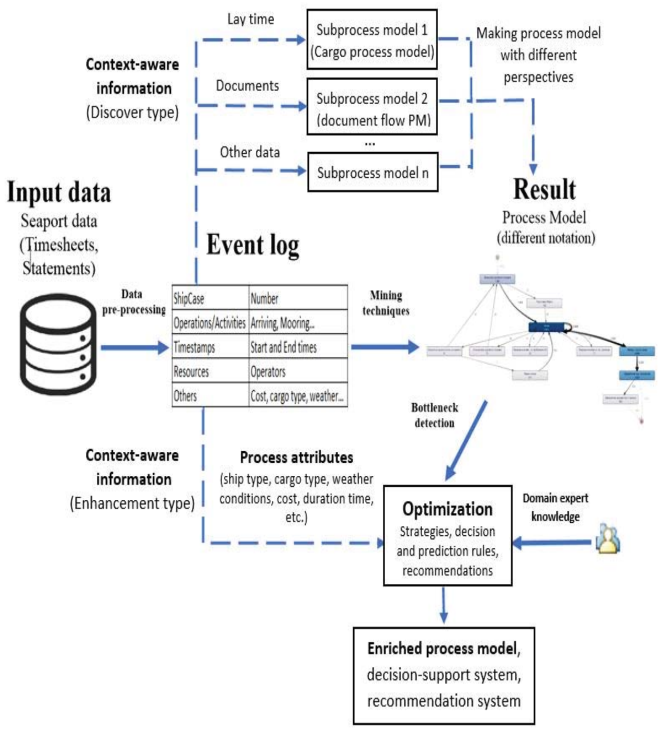


Fig. 2. Framework for acquiring an enriched ship-handling process model

Enhancement type: For this type of process mining, we can combine context-aware information with different data mining and machine learning techniques. Port terminals' activities have the multitude of information available behind the observed process. Type of cargo, type of ship, weather conditions, related operators, etc. – all these attributes have inner dependencies, which can be utilized for making decision rules, recommendation rules, and predictions [15]. Eventually, the main goal is to obtain an enriched process model with different perspectives in terms of a process map. The process model reflects a real process behavior (descriptive model) with the possibility of extension (scaling process), determination bottlenecks (traffic jam), detecting of deviations for the operational response, representation of different perspectives (control-flow, resources, performance).

## CONCLUSION AND DISCUSSION

Business processes are the core competence of enterprises. Improving business and operational processes is an essential task in logistics industries. In order to stay competitive in the market, flexibility is required. Generally, business process in logistics are designed beforehand. However, in real situations there are many factors that affect and change the regular processes to support the current conditions.

In this paper we discussed the application of process mining in the logistics domain, especially for seaport terminals. The real case of logistics in ship handling process at oil terminal was explored to evaluate the application of process mining. The obtained process map presents ship handling process from operation Arrival to operation Pilotage for leaving and helps to detect general process for different ship cases.

We integrated the concept of context awareness to enhance the performance of traditional process mining approach. The integration of context-aware information was illustrated in the proposed framework. The obtained process model can be used as a basis for making recommendation, decision-support and prediction systems. In the future work, the proposed framework will be extended to fit various types of logistics processes.


## ACKNOWLEDGMENT

This work was supported by The Ministry of Education, Youth and Sports of the Czech Republic from the National Programme of Sustainability (NPU II); project IT4Innovations excellence in science - LQ1602.


## REFERENCES


[1] R. S. Mans, M. H. Schonenberg, M. Song, W. M. van der Aalst, and P. J. Bakker, "Application of process mining in healthcare–a case study in a dutch hospital," in *International Joint Conference on Biomedical Engineering Systems and Technologies*, 2008, pp. 425–438.

[2] V. Rubin, C. W. Günther, W. M. Van Der Aalst, E. Kindler, B. F. Van Dongen, and W. Schäfer, "Process mining framework for software processes," in *International Conference on Software Process*, 2007, pp. 169–181.

[3] N. Trcˇka, M. Pechenizkiy, and W. van der Aalst, "Process mining from educational data," *Handb. Educ. Data Min.*, pp. 123–142, 2010.

[4] A. Rozinat and W. M. Van der Aalst, "Conformance checking of processes based on monitoring real behavior," *Inf. Syst.*, vol. 33, no. 1, pp. 64–95, 2008.

[5] S. J. Leemans, D. Fahland, and W. M. van der Aalst, "Process and Deviation Exploration with Inductive Visual Miner.," *BPM Demos*, vol. 1295, p. 46, 2014.

[6] G. M. Veiga and D. R. Ferreira, "Understanding spaghetti models with sequence clustering for ProM," in *International Conference on Business Process Management*, 2009, pp. 92–103.

[7] S. Goedertier, J. De Weerdt, D. Martens, J. Vanthienen, and B. Baesens, "Process discovery in event logs: An application in the telecom industry," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 1697–1710, 2011.

[8] A. Rozinat and W. M. van der Aalst, "Decision mining in ProM," in *International Conference on Business Process Management*, 2006, pp. 420–425.

[9] W. M. P. Van der Aalst, *Process mining: discovery, conformance and enhancement of business processes.* 2011. Springer-Verlag, 2008.

[10] A. K. A. de Medeiros, A. J. Weijters, and W. M. van der Aalst, "Genetic process mining: an experimental evaluation," *Data Min. Knowl. Discov.*, vol. 14, no. 2, pp. 245–304, 2007.

[11] C. W. Günther and W. M. Van Der Aalst, "Fuzzy mining–adaptive process simplification based on multi-perspective metrics," in *International conference on business process management*, 2007, pp. 328–343.

[12] Y. Wang, F. Caron, J. Vanthienen, L. Huang, and Y. Guo, "Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port," *Expert Syst. Appl.*, vol. 41, no. 1, pp. 195–209, 2014.

[13] J. Rudnitckaia, "Study of descriptive process model of ship handling at an oil terminal based on timesheets", *Vestnik of Astrakhan State Technical University*, no. 1, pp. 7–15, 2017.

[14] W. Intayoad and T. Becker, "Exploring the Relationship between Business Processes and Contextual Information in Manufacturing and Logistics Based on Event Logs," *Procedia CIRP*, vol. 72, pp. 557–562, 2018.

[15] J. Rudnitckaia, T. Hruska, "Prediction of times and delays for ship handling process based on a transition system" *Proceedings of the Int. Conf. on Harbor Maritime and Multimodal Logistics,* pp. 39–43, 2018.