Guided Process Discovery Approach According to Business Process Types

Lamghari Zineb[1], Saidi Rajaa[1,2], Radgui Maryam[1,2] and Rahmani Moulay Driss[1]
[1]*LRIT associated unit to CNRST (URAC 29) Rabat IT Center Faculty of Sciences, Mohammed V University in Rabat, Morocco*
[2]*SI2M Laboratory, National Institute of Statistics and Applied Economics, Morocco*

Keywords: Process mining, process discovery, business process type, process mining techniques, process model perspectives.

Abstract: Process discovery technique aims at automatically generating a process model that accurately describes a Business Process (BP) based on event data. Related discovery algorithms consider recorded events are only resulting from an operational BP type. While the management community defines three business process types, which are: Management, Support and Operational. They distinguish each BP type by different proprieties like the main business process objective as domain knowledge. This puts forward the lack of process discovery technique in obtaining process models according to business process types (Management and Support).

In this paper, we demonstrate that business process types can guide the process discovery technique in generating process models. A special interest is given to the use of process mining to deal with this challenge.

1 INTRODUCTION

Business Processes (BPs) are nowadays a crucial element in any organizational structure. They are established to manage and improve the company business. In this context, information systems assure the automation of BPs (Lamghari et al., 2021), by including Business Process Management (BPM) systems, Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), etc. In this sense, information systems record related event data to the BP execution, to analyse and guide issues concerning the creation of the company value. To achieve these objectives, process mining techniques are created.

Process mining is a scientific discipline that focuses on the analysis of event data logged during the execution of a BP, in order to discover, monitor and enhance BPs. It (Eck et al., 2015) consists of two major phases: Pre-processing and Processing. Depending on the process subject, we extract data from Repository. Once the data has been extracted; we proceed to filter it from deficiencies in the log cleaning step. By doing so, we arrive to the process model discovery technique. Next, we apply the conformance checking step, which aims at evaluating how well this discovered process model corresponds to reality. This gives input for changing parameters in the main process, and requires going back to the data cleaning step, to re-apply the discovery technique. After evaluation, we diagnose all obtained results, to provide inputs for business process improvements. Therefore, event data analysis and the discovery technique are required for obtaining suitable process models that takes into consideration: event logs content, their levels of abstraction (Aalst, 2016) and the process model perspectives (Bogarin et al., 2018). By necessity, a suitable process model reflects the business reality and met requirements.

Process model perspectives are: Control-flow (The order in which its activities should be executed), Organizational (The resources required for the execution of a process and how they interact with each other), data-flow (The data objects created and updated during the execution of the process) and time (The time-related aspects of the process). In this work, we assume that the pre-processing phase works directly on cleaning event logs, while the processing phase focuses on analysing events and constructing process models.

In this sense, depending on the input data and the questions that need to be answered, a suitable process model abstraction can be presented. The model may be too abstract and thus unable to answer relevant questions. The model may also be too detailed, e.g., the required input cannot be obtained or the model becomes too complex to be fully understood.In this context, many process discovery methods assume that recorded events correspond to meaningful activities in the instances of a process. However, events may be recorded on different levels of granularity. Some events may refer to activities on a high level of abstraction. Their execution is easily recognizable for process workers. Other events may be recorded with a lower level of abstraction. Multiple of such low-level events may referee to a recognizable high-level activity. When discovering processes based on those low-level events, the resulting process model can impact process workers structure. Consequently, the discovered model represents the wrong level of abstraction. Therefore, the event data Domain Knowledge (DK) is the crucial parameter that can guide the process discovery technique ❶, in terms of diagnosing and representing process models according to their DK. Besides, the management community has defined the BP type as the DK that can impact (Harmon, 2015; Lamghari et al., 2018) the BP representation and treatment ❷.

Based on these two information ❶ and ❷, we assume the following hypothesis: BP type as DK may influence process model that can be discovered with process discovery techniques, in terms of perspectives priority, i.e., according to DK which perspective will be treated or combined firstly with the control-flow perspective for guiding the process discovery in generating process models.

The BP types (Burattin, 2015) are:

- The Management BP: Describes the process of the product or services realization that is provided by the company to their customers.
- The Operational BP: Defines the organization strategy.
- The Support BP: presents the process that offers resources to other processes to ensure the smooth running of the company.

Thus, we need an approach that can tackle the treatment of events and discover models according to BP types (Burattin, 2015), towards guiding the process discovery technique in discovering suitable models. This can be achieved using DK as the BP type related to multiple process model perspectives.

In this context, we must discuss different related issues that involve the intersection between BP types and process mining. For this purpose, our paper proposes an approach that demonstrates how BP type, as DK, will guide and impact the process discovery process. In this sense, our paper is organized as follows: Section 2 presents the still encountered related issues of the intersection between BP types and process mining. The section 3 details our approach that consists of guiding the process discovery technique according to BP types. Conclusion and further directions are presented in section 4.

2 BRIDGING THE GAP BETWEEN BUSINESS PROCESS TYPES AND PROCESS MINING

In this section, we present the still encountered issues related to the intersection between BP types and process mining, which are: Event data quality, Management view (Configuration phase, BP status, and BP execution), Levels of representing process models or their abstraction that includes process model perspectives and the clustering technique.

2.1 Event Data Quality

The quality is achieved by filtering out noise (incorrectly logged), incomplete (missing events), chaotic (arbitrary executed) and infrequent behaviours from event data. These deficiencies are denoted, in Fig.1, respectively as (Domain_knowledge_N, (Domain_knowledge_I, etc).

In the literature, we find many research papers, like (Conforti et al., 2017; s. Suriadi et al., 2017), dealing with this topic. This issue has received a significant focus from the research community, where all related ambiguities have been resolved. Here, we propose to use the BP type as an additional filtering parameter, to refine the main BP context.

2.2 Management View

Information systems record event data. Their configuration is one of the most prominent tasks related to event data preparation.

Generally, a BP is defined with specific parameters as DK (BP type), BP status (informs about the BP lifecycle) and other proprieties that depend on the BP objective and the organization requirements. These parameters describe the configuration file content, to record behaviours and avoid errors that may be provided during the BP execution. The point has not matched, in this regard, is the definition of the BP type, its particularity and how it may impact the process discovery technique and the process model per-

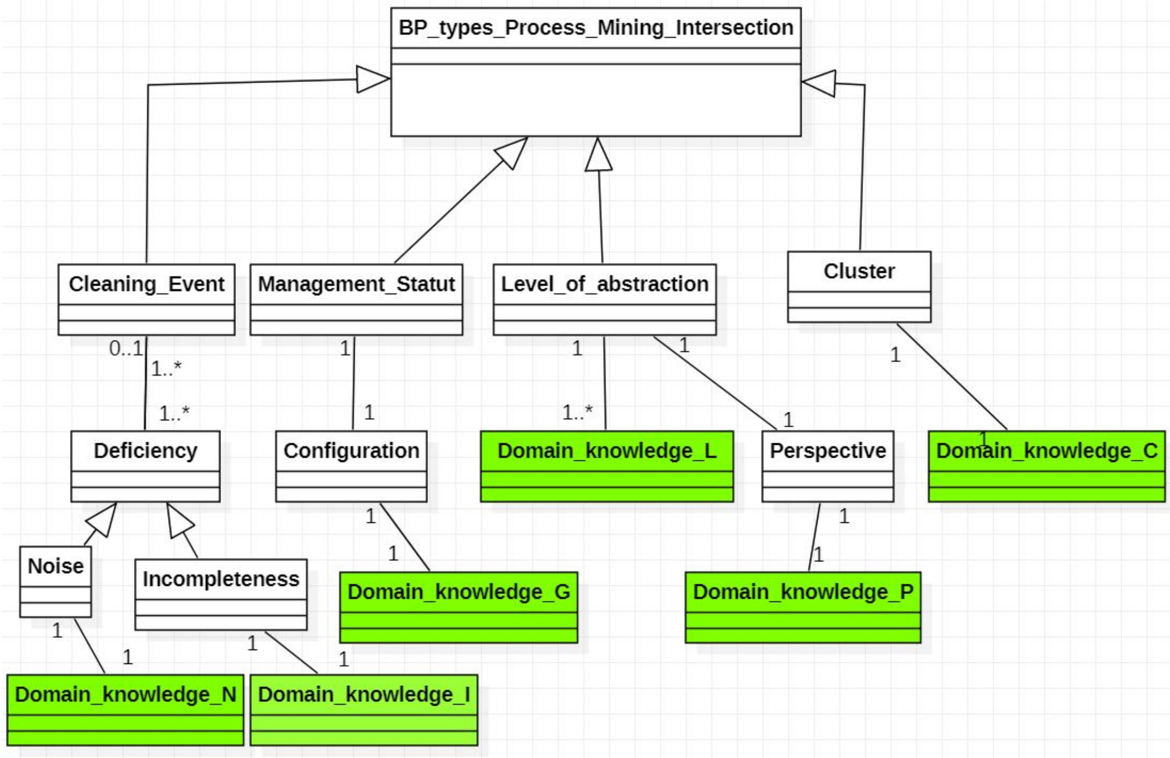

Figure 1: Related issues to the intersection between BP types and process mining.

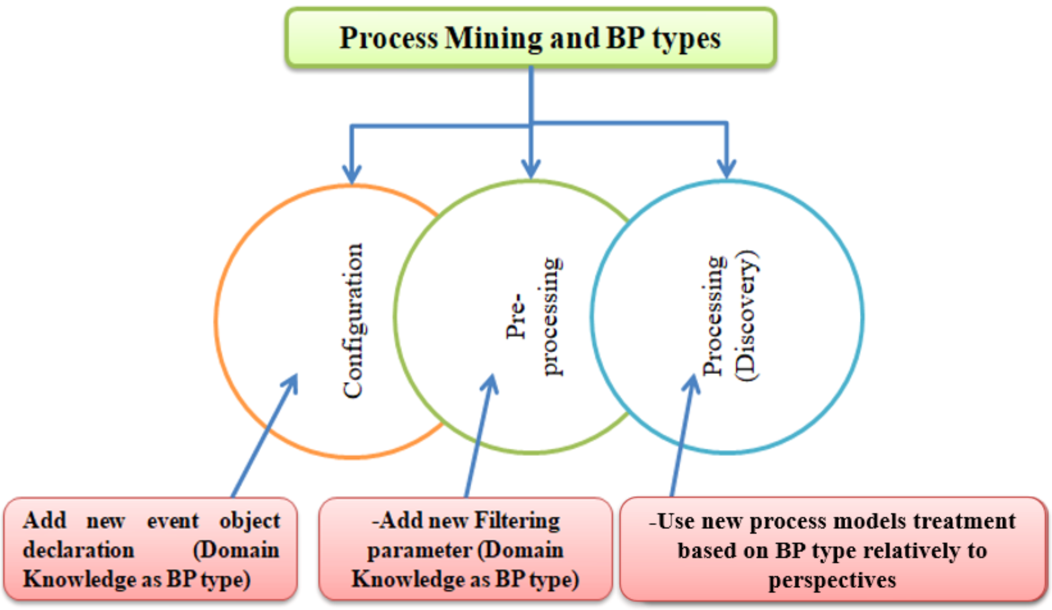

Figure 2: Our contribution aspects.

spectives, in the term of representation. Available research (Bogarin et al., 2018; Conforti et al., 2017; Boubaker et al., 2016) consider recorded events that are resulting from an operational BP type. These studies do not mention the two other BP types (Support and Management), while the BP type can influence the representation of process models that can be discovered with process discovery techniques. In this respect, the DK or BP type is implicitly defined in the configuration file, it has not been declared as attribute in the event object (a part of the configuration file). This provides another untreated point to consider. Thus, the propriety we propose to use is to declare BP type as an attribute that will determine the BP domain knowledge.

2.3 Levels of Abstraction

The process discovery technique generates models using the 2-D method (Aalst, 2016), where the process model can be viewed in different levels of abstraction. As mentioned in the (Bogarin et al., 2018) events of low-level are mapped to activity instances of high-level, based on specific DK. The correspondence between these two levels is treated and provided adequate results (Bogarin et al., 2018; Baier et al., 2016; Ciccio et al., 2018). For instance, one high-level activity instance may result in multiple low-level events being recorded and, vice versa, one such low-level event may relate to multiple high-level activity instances.

The point not yet treated is the BP type (Domain_knowledge_L in Fig.1) impact on the process model abstraction, by necessity process model perspectives' priority, i.e.; according to the BP type as DK, which process model perspective will be combined with the control-flow perspective.

2.4 Clustering Technique

The clustering technique is treated in the processing phase as mentioned in (Oliveira and Queiroz, 2020). We do not observe it in the pre-processing phase. Introducing this technique into the pre-processing phase can refine the final process model construction. Moreover, this technique based on DK as BP type definition that groups a set of activity instances with the same context (BP type), to determine a clustered process model. In this sense, for event logs sequences, we can determine the set of similar data (cluster).

2.5 Synthesis

According to this discussion, we have detected some untreated stages related to BP types (Cf. Fig.1). These questionable stages emerged new issues, relatively to process mining. We have also explored an additional treated phase, which is the configuration phase. Therefore, our contribution will focus on three phases (configuration, pre-processing and processing). In this sense, DK is our guided approach parameter. Thus, we will define DK as (Cf. Fig.2 and Fig.1):

- BP type presents event object in the configuration phase (Domain_knowledge_G).
- BP type presents an additional filter to define the main cluster for pre-processing phase (Domain_knowledge_C).
- BP type as BP type objective and related to process model perspectives (Domain_knowledge_G).

3 OUR GUIDED PROCESS DISCOVERY APPROACH

Our approach aims at guiding the process discovery technique in representing process models according to different BP types. It consists of three phases: configuration, pre-processing and processing (Cf. Fig.3).

Our approach process starts by configuring the information system, where all BPs will be executed and recorded according to the event object parameters. Once, the selected BP is executed, we verify if the configuration elements are successfully achieved. If this later is well done, we proceed to the pre-processing phase, which aims at cleaning event data from deficiencies. If not, we loop back to the configuration phase. Finally, we pass to the processing phase, to obtain process model representation.
Throughout our approach phases:

- We add a new event object in the configuration phase ①. After executing a BP ② and collecting recorded event logs, we arrive to the second operation named ③ that uses an additional filtering attribute, which is the BP type of the pre-processing phase. Finally, we apply ④ the processing phase techniques for mining process models according to BP types, by necessity the BP type objective and the process model perspectives (control-flow, time, organizational).
- We define different DK, and we obtain different outputs (1-main frame based on the BP type, 2-main cluster based on the BP type, 3-clustered process model based on the BP type).

3.1 Configuration Phase

The configuration phase (Cf. Fig.3-①) takes as an input a BP model. It gives idea and insights about the BP type. Indeed, the configuration phase takes into consideration the BP type as DK.
The reason to choose the BP type as DK within the configuration phase is:

- The use of management view from early stage.

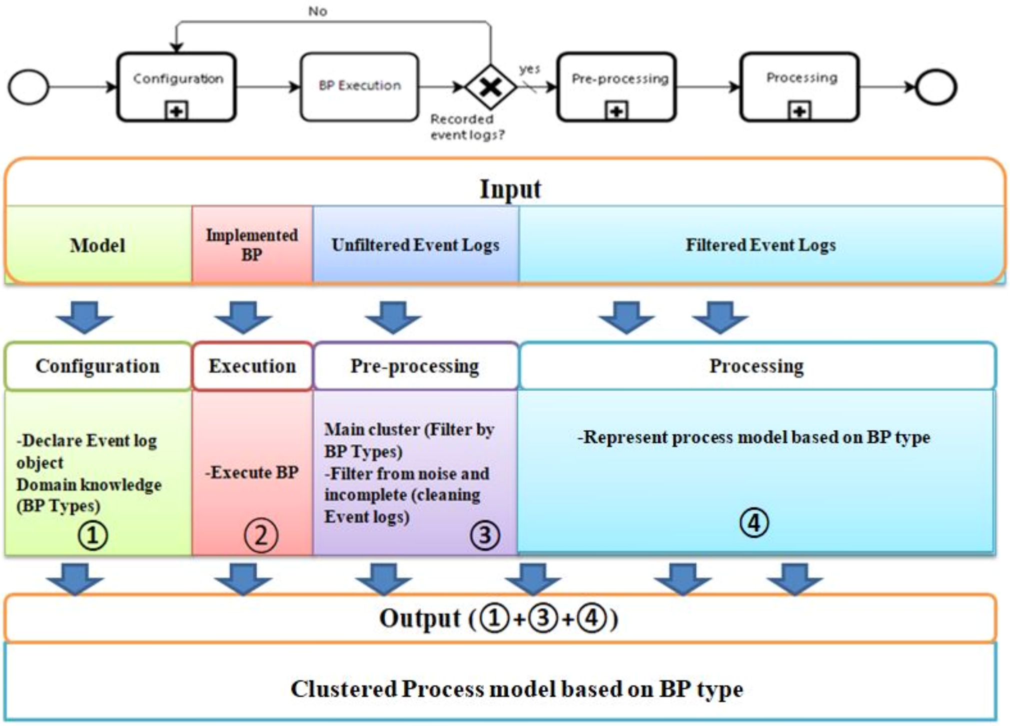

Figure 3: Guided Process Discovery approach phases.

- The BP type impacts on the next two phases.
- Preparing events in the configuration phase as an early basic filter.

The configuration process consists of:

- Determining the BP domain knowledge as the operational BP, the management BP, and the support BP.
- Define the event object content.
- Declare the BP type as attribute in the event object.
- Activate the recording option.

3.2 Execution Phase

During the execution phase (Cf. Fig.3-②), we record BP event logs. In this context, the minimum information required of a standard format for an event log is: The case ID is a unique identifier for a process instance, the information stated in chronological order according to events. Also, additional information is possible, we find attributes such as timestamp (the time when the activity took place), resources (who performed the activity), transaction type and costs associated with the event.

3.3 Pre-processing Phase

After executing the BP in (Cf. Fig.3-③), the pre-processing process (Cf. Fig.3-③) starts by extracting events in the adequate form, to apply the filtering operation. This operation cleans events from deficiencies: noise, incompleteness, chaotic and infrequent behaviours. After obtaining a cleaned event logs, we define the main cluster using the BP type parameter (the main frame of event data). The reason for introducing a main cluster, in the pre-processing phase, is to provide an advanced recognition technique using the BP type.

3.4 Processing

After declaring the BP type attribute in the configuration phase and defining the main cluster based on the BP type in the pre-processing phase, we arrive to the processing phase (Cf. Fig.3-④). Therefore, the processing phase consists of:

- Using the clustering technique, to get a clustered process model, i.e., clustered process model based on the BP type. We define clusters semantically (Staab and Studer, 2010).
- Defining the process model perspectives priority: The BP type gives idea on the main BP objective

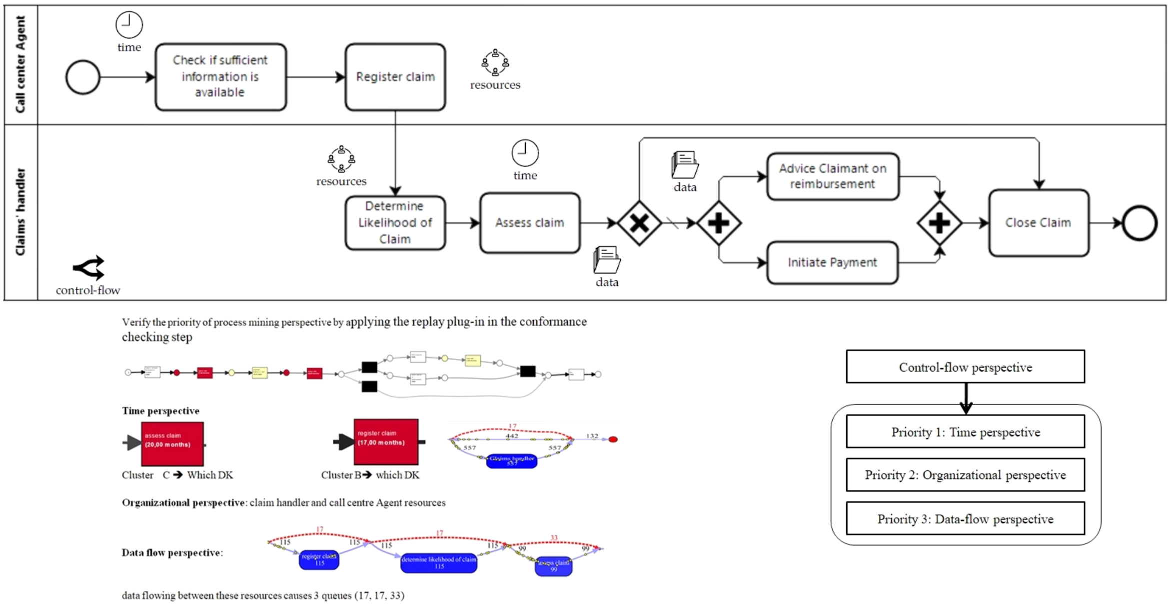

Figure 4: Example of process model perspectives order.

and gives insights on the process model perspectives order (In which order perspectives will be combined with the control-flow) within the logic of:
The BP type + the BP objective → Results the order of representing process model details:

- Operational BP + Satisfaction → (1. Time, 2. Resource, 3. Data) Satisfy the client means that a process respects time limitation. If not, I will verify why resource did not respect the time condition in order to the data-flow into this BP.
- Management BP + Management method → (1. Data, 2. Resource, 3. Time) Achieve the management method objective in respect to the data-flow used, into this BP, by resources in time condition.
- Support BP + Resource Support → (1. Resource, 2. Time, 3. Data-flow) Obtain resources support related to the time indication and the data-flow of this BP.

The perspectives order can be verified in the configuration phase, and it can provide more insights on the probably detected deviations.

3.5 Illustrative Example

In this sub-section, we treat an operational BP about claim's handler `http://www.processmining.org/event_logs_and_models_used_in_book`. We aim to verify our process model perspectives order proposition.

In our example, the claim's handler BP makes sure that claims are handled efficiently and that payment for valid claims is made. Also, this process consists of making decisions on the extent and validity of a claim, and the checking for any potentially fraudulent activity (Cf. Fig.4).
To represent our process model example, we use the heuristic miner algorithm (Weijters et al., 2006). By necessity, this algorithm takes into consideration process quality metrics: fitness, precision, generalization, and simplicity.
According to this BP type, we can firstly detect the time deviation. Then, by defining on which activity the deviation is provided, we can define responsible resources for this deviation and respectively data-flow contributed to this deviation. Therefore, we observe different advantages of our proposed approach, comparing to the ordinal process mining project phases (Eck et al., 2015):

- BP type as domain knowledge used within three phases.
- Clustering technique applied from the preprocessing phase.
- Filtering insights mentioned from the configuration phase.
- Conformance checking indications appeared from the discovery step.
- Basic filtering applied from the configuration phase (By BP types).
- Combining the management view with the pro-

cess model perspectives, which are resulting the process model perspectives priority concept.

- New process model representation according to BP types.

3.6 Synthesis

Our approach is applied on three phases: configuration, pre-processing and processing. Throughout these phases, we have presented a guided process discovery approach by BP types. The general idea consists of taking into consideration the impact of BP types on process model representation.
We have treated the still encountered related issues of the intersection between BP types and process mining: Management view, Clustering technique and Process model. The management view takes into consideration the impact of the BP type on event logs recognition. The clustering technique uses BP type as context, to group a set of commune activity instances. The process model perspectives propose a flow of representation guided by BP types. In this sense, the conformance checking technique can approve the application order of process model perspectives according to BP types.

4 CONCLUSIONS

In this paper, we present an approach dealing with the process discovery technique according to BP types. Indeed, we aim to guide the process mining discovery technique, in order to generate suitable process model for each BP type. For this purpose, we investigate the still encountered issues related to the intersection between BP types and process mining. We observe four mains required objectives: the management view, process model perspectives and the clustering technique. In this context, we match each challenge with a specific phase of our proposed approach. Consequently, our approach is applied on three phases: configuration, pre-processing and processing. In this respect, the configuration phase declares BP type as event object, to define selected event data (Lamghari et al., 2019) by BP type. Then, the pre-processing phase use a new filter, which aims to refine event data frame by BP type. Last, the processing phase treats event logs, using the correspondence between BP types and process model perspectives' priority, to represent process models according to BP types. This helps in acquiring insights on which order perspectives could be combined to the control-flow perspective. As further research, we plan to develop a full plug-in that can be implemented into the Prom tool, for applying our proposed guided process discovery approach according to BP types and improvement metrics (Lamghari et al., 2019).

ACKNOWLEDGEMENTS

This work was supported by the National Center for Scientific and Technical Research (CNRST) in Rabat, Morocco.

REFERENCES

Aalst, W. M. P. V. D. (2016). *Data science in action*. Springer, 2nd edition.

Baier, T., Ciccio, C. D., Mendling, J., and Weske, M. (2016). Matching events and activities by integrating behavioral aspects and label analysis. *International Journal of Business Process Integration and Management*, 17(2):573–598.

Bogarin, A., Cerezo, R., and Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(1):12–30.

Boubaker, S., Mammar, A., Graiet, M., and Gaaboul, W. (2016). A formal guidance approach for correct process configuration. In *International Conference on Service-Oriented Computing*. Springer.

Burattin, A. (2015). *Process Mining Techniques in Business Environments*. Springer.

Ciccio, C. D., Maggi, J., Montali, M., and Mendling (2018). Matching events and activities by integrating behavioral aspects and label analysis. *Information Systems*, 78(2):144–161.

Conforti, R., Rosa, M. L., and Hofstede, H. T. (2017). Filtering out infrequent behavior from business process event logs. *IEEE Transactions on Knowledge and Data Engineering*, 19(2):300–314.

Eck, M. V., Lu, X., Leemans, S., and aalst, W. D. (2015). Pm2: A process mining project methodology. In *The 27th International Conference on Advanced Information Systems Engineering Proceedings*. Springer.

Harmon, P. (2015). *Business process change: a business process management guide for managers and process professionals*. Morgan Kaufmann.

Lamghari, Z., Radgui, M., Saidi, R., and Rahmani, M. (2018). A set of indicators for bpm life cycle improvement. In *International Conference on Intelligent Systems and Computer Vision*. IEEE.

Lamghari, Z., Radgui, M., Saidi, R., and Rahmani, M. (2019). Defining business process improvement metrics based on bpm life cycle and process mining techniques. *International Journal of Business Process Integration and Management*, 9(2):107–133.

Lamghari, Z., Saidi, R., Radgui, M., and Rahmani, M. (2021). An operational support approach for mining unstructured business processes. *Revista de Informática Teórica e Aplicada*, 28(1):23–38.

Oliveira, M. D. S. and Queiroz, S. (2020). Unsupervised feature selection methodology for clustering in high dimensionality datasets. *Revista de Informática Teórica e Aplicada*, 27(2):30–41.

s. Suriadi, Andrews, R., and Hofstede, H. T. (2017). Filtering out infrequent behavior from business process event logs. *Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs*, 64(2):132–150.

Staab, S. and Studer, R. (2010). *Handbook on ontologies*. pringer Science and Business Media.

Weijters, A., Aalst, W. D., and Medeiros, A. D. (2006). *Process mining with the heuristics miner-algorithm*. Technische Universiteit Eindhoven.