

# Abstracting Low-level Event Data for Meaningful Process Analysis

Adrian Rebmann<sup>1</sup>

<sup>1</sup>*Data and Web Science Group, University of Mannheim, Mannheim, Germany*

## Abstract

Most process mining techniques assume the event data to be on the right level of detail for analysis, i.e. the level of activities. Low-level recording and a high flexibility of the underlying process lead to results of limited value, when applying process discovery techniques, as output models are overly complex. In this PhD project, we aim to develop approaches for event abstraction that balance between reducing the complexity of process mining results, while ensuring meaningful analyses by incorporating previously disregarded perspectives. Specifically, we will focus on three research streams, developing solutions that consider the meaning of events, take the purpose of analyses into account, and address online settings. In this paper, we outline the current state and future plans for each of these research streams.

## Keywords

Process analysis, Event abstraction, Semantic labeling, Purpose-driven abstraction, Stream-based process mining

## 1. Introduction

Process mining enables the analysis of processes based on sequences of events recorded by information systems. This leads to actionable insights into how a process is really executed. The majority of the available process mining techniques still makes the assumption that event data are captured on the level of activities relevant for analysis.

However, this is often not the case, e.g., when events are recorded on the level of clicks in a system. Besides fine-granular event recording, also mixed-granular recording and an inherent flexibility of the underlying process leads to high variability in the event sequences [1]. Applying process discovery techniques to such data can lead to results that are hard to interpret due to a high complexity of the output models. To tackle this issue, event abstraction approaches have been proposed recently [2, 1]. While such approaches incorporate different perspectives, none of them actually considers the meaning of events, e.g., by taking into account the specific actions applied to certain objects in a process instance. This can be problematic, since the meaning can be crucial for the decision whether or not to group events. For instance, consider two events “check insurance” and “draw blood sample” that always occur close to each other in the traces of an event log. If only the control flow is considered, these are prime candidates

---


*Proceedings of the Demonstration & Resources Track, Best BPM Dissertation Award, and Doctoral Consortium at BPM 2021 co-located with the 19th International Conference on Business Process Management, BPM 2021, Rome, Italy, September 6-10, 2021*

✉ [rebmann@informatik.uni-mannheim.de](mailto:rebmann@informatik.uni-mannheim.de) (A. Rebmann)

ORCID [0000-0001-7009-4637](https://orcid.org/0000-0001-7009-4637) (A. Rebmann)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

to be grouped. However, from a semantic perspective grouping them into the same high-level activity is undesirable as they are clearly different activities. A further problem is that users of current event abstraction approaches have no control over the characteristics of the output event log. This is necessary to account for the specific purpose of a user's process analysis goal or to incorporate their knowledge about high-level activities. For instance, a user that aims to analyze the resource perspective does not want low-level events to be grouped, if these were performed by different resources. Finally, especially in online settings event data is typically recorded on a low-level, in the form of sensor readings or click stream data, for instance. This calls for online abstraction techniques to reduce complexity and data volume in a timely manner, to enable the application of stream-based process mining techniques, such as online process discovery [3]. However, there is limited work on online event abstraction in process mining [1].

In this PhD project, we aim to overcome these gaps. First, we aim to use the meaning of events extracted from textual payload of events to guide abstraction techniques towards better results. Moreover, we consider characteristics to which abstracted event sequences should adhere. These can be defined by a user of the technique, as they know best what the purpose of their overall analysis is. This allows for more meaningful, purpose-driven abstractions. Finally, we will focus on online scenarios, developing approaches to group events in real-time.

The next section discusses the state of the art of event abstraction in process mining. Section 3 gives an overview of the three research streams addressed in this project, their current state, and the respective research plans. Section 4 concludes the paper.

## 2. State of the Art

Two recent literature reviews give a comprehensive overview of the state of the art in event abstraction in process mining, which indicates an increasing interest in the topic [2, 1]. As shown there, the assumptions these approaches make differ greatly. The most prominent assumption is that the target level of abstraction is known (cf. [4, 5]). This knowledge is assumed to come from existing process models, or from a set of event sequences on two levels of which a mapping is known. This, however, can only be assumed for a fraction of real-life scenarios. Thus, unsupervised or semi-supervised event abstraction approaches have been proposed (cf. [6, 7]). In this project, we focus on such unsupervised and semi-supervised approaches and consider event abstraction as the meaningful grouping of low-level events into higher-level activities.

In a related approach the authors suggest to use semantic similarity between event labels to aggregate events. However, the actual event abstraction is not described [8]. Regarding purpose-driven event abstraction, the most closely related approach was developed by Mannhardt et al. [9], which takes as input the activities defined as low-level process models. This requires a lot of effort from the user and, most importantly, very detailed knowledge of how activities manifest themselves in low-level event sequences. With respect to event abstraction in streaming scenarios, related approaches can be found in work that combines process mining with online activity recognition [10]. Moreover, Complex Event Processing (CEP) techniques offer concepts for event abstraction in process mining [11].

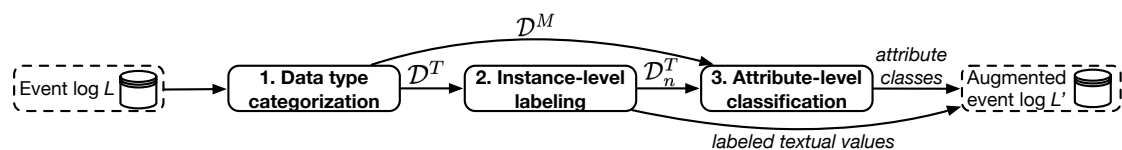
### 3. Current State and Research Plans

In this PhD project we aim to develop solutions to event abstraction in three research streams. We (1) consider the meaning of events, (2) take the analysis purpose into account, and (3) address online and offline scenarios with their own specific challenges. We follow a design science paradigm [12] throughout the project. The objective of a design-oriented approach is the development of artifacts that address relevant, unresolved problems. The relevance of the artifacts to be developed is derived from the increasing complexity of processes in organizations, their increasing flexibility, and the increasing level of detail of recorded event data. This calls for solutions to reduce the complexity of event data in offline as well as online settings.

For the first two research streams, initial concepts and approaches have already been developed, while the third will be addressed in future work. In the following, we give an overview of the current state and the plans for each of the research streams.

#### 3.1. Using semantic process information for event abstraction

Events frequently contain textual information that allows to reason about the meaning of underlying actions, objects, and other semantic components in the process. These can contain valuable information of the level of abstraction and related events that can be used for abstraction. Our work on extracting semantic components from the events recorded in event logs yielded a first publication in this project [13]. The goal here is to label the values of event attributes with semantic roles. To achieve this, our approach follows three main steps, visualized in Figure 1.



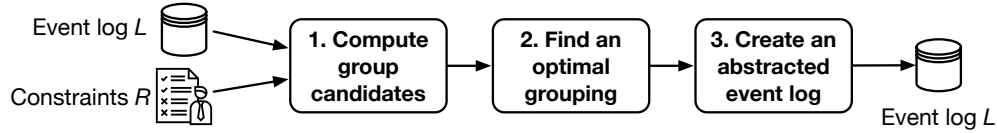
**Figure 1:** Main steps of the semantic extraction approach.

The outcome of developed approach can already be used for event abstraction tasks. For instance, we can group together multiple actions applied to a business object, lifting the log to an object-centric perspective. However, we plan to extend this work to be able to group low-level events into *meaningful stages* in a process using the meaning of the extracted components and how they behave throughout a process instance. Furthermore, once events are grouped, there is a need for a meaningful name for this higher-level event. Thus, we aim to develop an approach to automatically derive such names taking into account behavioral relations in the process as well as extracted semantic information. We plan to evaluate these techniques based on publicly available real-life event logs, which were shown to contain diverse natural language labels [13]. Several also contain information about higher levels of the respective process, e.g., attributes indicating the subprocess, which can be used to validate our results.

#### 3.2. Purpose-driven event log abstraction

We also aim to account for the specific knowledge a user can bring to an event log abstraction task and consider the purpose of subsequent process analyses. For instance, a user who wants to

analyze an event log with respect to the resource perspective wants to be sure that each activity in the abstracted log was performed by a single actor per case. This should be incorporated by an event abstraction approach, leading to more purpose-oriented event logs. Hence, the goal is to find an optimal grouping of a set of event classes, while user-defined constraints are met. These constraints can be defined in a declarative manner on the entire grouping, the high-level activities, and specific activity instances including their attributes in the abstracted log. We are currently developing an approach that achieves this following the steps visualized in Figure 2.



**Figure 2:** Main steps of the constraint-based event log abstraction approach.

We evaluate this approach, by employing it in case studies using real-life logs to show the value of considering constraints in event abstraction. Furthermore, we assess its abstraction quality based on the complexity reduction of process models discovered after abstraction.

Using this approach, the user can then incorporate their knowledge, respectively their idea of characteristics to which an activity should adhere. Thus, event groups become more meaningful and produce more understandable results through just the right amount of abstraction. This contributes to the overall goal of balancing simplicity and utility of abstracted event data.

### 3.3. Event abstraction over event streams

The third scenario shifts the focus from post hoc to online event abstraction. The challenge here is the sheer volume and velocity of events arriving, which calls for efficient single pass solutions for event abstraction. Therefore, in this third research stream we will build on the results of the previous ones to develop techniques for event abstraction in streaming settings. Existing work on stream-based process mining, such as online conformance checking [14, 15] and online concept drift detection [16, 17] can serve as a starting point. An initial literature study was conducted for this research stream, which revealed the potential of adopting ideas from other fields. Especially CEP [11] and stream-based sequential pattern mining [18] can serve as a foundation on which to build event abstraction techniques for online settings.

As a first step we aim to apply parts of our purpose-driven abstraction approach, as constraints can be used to guide event pattern detection in a stream. Yet, the computational complexity of mining such patterns calls for approximation techniques to make such an approach feasible.

## 4. Conclusion

We gave an overview of the current state and the research plans of the proposed PhD project. Therein, we develop solutions to event abstraction that consider the meaning of events and the purpose of subsequent process analyses. Moreover, both offline and online settings are addressed. All developed concepts and solutions serve the overarching goal of enabling more meaningful process analyses through event abstraction.

## References

- [1] S. J. van Zelst, F. Mannhardt, M. de Leoni, A. Koschmider, Event abstraction in process mining: literature review and taxonomy, *Granular Computing* 2 (2020).
- [2] K. Diba, K. Batoulis, M. Weidlich, M. Weske, Extraction, correlation, and abstraction of event data for process mining, *Wiley Interdisciplinary Reviews* 10 (2020) 1–31.
- [3] S. J. van Zelst, B. F. van Dongen, W. M. van der Aalst, Event stream-based process discovery using abstract representations, *Knowledge and Information Systems* 54 (2018) 407–435.
- [4] T. Baier, J. Mendling, M. Weske, Bridging abstraction layers in process mining, *Information Systems* 46 (2014) 123–139.
- [5] N. Tax, N. Sidorova, R. Haakma, W. M. van der Aalst, Event abstraction for process mining using supervised learning techniques, in: *Proceedings of SAI*, Springer, 2016, pp. 251–269.
- [6] M. De Leoni, S. Dünder, Event-log abstraction using batch session identification and clustering, *Proceedings of the ACM Symposium on Applied Computing* (2020) 36–44.
- [7] C. W. Gunther, A. Rozinat, W. M. Van Der Aalst, Activity mining by global trace segmentation, *Lecture Notes in Business Information Processing* 43 (2010) 128–139.
- [8] P. H. P. Richetti, F. A. Baião, F. M. Santoro, Declarative process mining: Reducing discovered models complexity by pre-processing event logs, in: S. Sadiq, P. Soffer, H. Völzer (Eds.), *BPM*, Springer, Cham, 2014, pp. 400–407.
- [9] F. Mannhardt, M. D. Leoni, H. A. Reijers, M. D. Leoni, H. A. Reijers, Guided process discovery—a pattern-based approach, *Information systems* 76 (2018) 1–18.
- [10] F. Mannhardt, R. Bovo, M. Oliveira, S. Julier, A taxonomy for combining activity recognition and process discovery in industrial environments, in: H. Yin, D. Camacho, P. Novais, A. J. Tallón-Ballesteros (Eds.), *IDEAL 2018*, Springer, Cham, 2018, pp. 84–93.
- [11] P. Soffer, A. Hinze, A. Koschmider, H. Ziekow, C. Di Ciccio, B. Koldehofe, O. Kopp, A. Jacobsen, J. Sürmeli, W. Song, From event streams to process models and back: Challenges and opportunities, *Information Systems* 81 (2019) 181–200.
- [12] A. R. Hevner, S. T. March, J. Park, S. Ram, Design science in information systems research, *MIS quarterly* (2004) 75–105.
- [13] A. Rebmann, H. van der Aa, Extracting semantic process information from the natural language in event logs, in: *CAiSE*, Springer, 2021, pp. 57–74.
- [14] A. Burattin, S. J. van Zelst, A. Armas-Cervantes, B. F. van Dongen, J. Carmona, Online conformance checking using behavioural patterns, in: *BPM*, Springer, 2018, pp. 250–267.
- [15] D. Schuster, S. J. van Zelst, Online process monitoring using incremental state-space expansion: An exact algorithm, in: D. Fahland, C. Ghidini, J. Becker, M. Dumas (Eds.), *BPM*, Springer, Cham, 2020, pp. 147–164.
- [16] M. Hassani, Concept drift detection of event streams using an adaptive window., in: *ECMS*, 2019, pp. 230–239.
- [17] F. Stertz, S. Rinderle-Ma, J. Mangler, Analyzing process concept drifts based on sensor event streams during runtime, in: D. Fahland, C. Ghidini, J. Becker, M. Dumas (Eds.), *BPM*, Springer, Cham, 2020, pp. 202–219.
- [18] M. Hassani, S. J. van Zelst, W. M. van der Aalst, On the application of sequential pattern mining primitives to process discovery: Overview, outlook and opportunity identification, *Wiley interdisciplinary reviews: Data Mining and Knowledge Discovery* 9 (2019).