

## Domain-specific Event Abstraction

Finn Klessascheck<sup>1</sup>, Tom Lichtenstein<sup>1</sup>, Martin Meier<sup>1</sup>, Simon Remy<sup>1</sup>, Jan Philipp Sachs<sup>2,4</sup>, Luise Pufahl<sup>3</sup>, Riccardo Miotto<sup>4,5</sup>, Erwin Böttinger<sup>2,4</sup>, and Mathias Weske<sup>1</sup>

<sup>1</sup>Hasso Plattner Institute (HPI), University of Potsdam, Potsdam, Germany

<sup>2</sup>Digital Health Center, HPI, University of Potsdam, Potsdam, Germany

<sup>3</sup>Software & Business Engineering, Technische Universität Berlin, Berlin, Germany

<sup>4</sup>Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, USA

<sup>5</sup>Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, USA

**Abstract.** Process mining aims at deriving process knowledge from event logs, which contain data recorded during process executions. Typically, event logs need to be generated from process execution data, stored in different kinds of information systems. In complex domains like healthcare, data is available only at different levels of granularity. Event abstraction techniques allow the transformation of events to a common level of granularity, which enables effective process mining. Existing event abstraction techniques do not sufficiently take into account domain knowledge and, as a result, fail to deliver suitable event logs in complex application domains. This paper presents an event abstraction method based on domain ontologies. We show that the method introduced generates semantically meaningful high-level events, suitable for process mining; it is evaluated on real-world patient treatment data of a large U.S. health system.

**Keywords:** Process mining, Event abstraction, Domain knowledge, Healthcare

## 1 Introduction

Many organizations have an inherent interest to monitor and understand their processes. For example, analyzing and adopting processes can improve their overall efficiency, ensure that legal requirements are met, and maintain a desired quality level. To this end, process mining provides techniques to analyze processes based on event data recorded during their execution. However, such event data is not always available in the necessary format and often differs in its granularity in complex settings. This also applies to treatment processes in hospitals, which are typically highly heterogeneous, complex, multidisciplinary, ad-hoc, and susceptible to change [1]. In the past, process mining has been proven as a technique well-suited to derive an understanding of medical processes, like patient-flows, and to improve them accordingly [2].

When extracting event data for process mining of electronic health records (EHRs), multiple data sources have to be tapped into, including hospital information systems. This variety of data sources and differences in data granularity leads to a mismatch in the level of abstraction between different events. Moreover, the fact that many events are documented manually by physicians or other medical personnel typically leads to varying degrees of detail in the recorded events. Using the resulting event logs would result in complex process models [3]–[5]. In order to generate event logs with events of comparable granularity and to elicit useful process models, event abstraction is needed. To this end, a rich set of research works on event abstraction



















- [5] X. Lu, A. Gal, and H. A. Reijers, "Discovering hierarchical processes using flexible activity trees for event abstraction," in *2nd International Conference on Process Mining, ICPM*, 2020, pp. 145–152.
- [6] K. Diba, K. Batoulis, M. Weidlich, and M. Weske, "Extraction, correlation, and abstraction of event data for process mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 3, 2020.
- [7] A. Qaseem, T. J. Wilt, R. M. McLean, and M. A. Forciea, "Noninvasive treatments for acute, subacute, and chronic low back pain: A clinical practice guideline from the american college of physicians," *Annals of Internal Medicine*, vol. 166, no. 7, pp. 514–530, 2017.
- [8] F. Folino, M. Guarascio, and L. Pontieri, "Mining predictive process models out of low-level multidimensional logs," in *Advanced Information Systems Engineering (CAiSE)*, ser. LNCS, vol. 8484, Springer, 2014, pp. 533–547.
- [9] P. H. P. Richetti, F. A. Baião, and F. M. Santoro, "Declarative process mining: Reducing discovered models complexity by pre-processing event logs," in *Business Process Management BPM*, ser. LNCS, vol. 8659, Springer, 2014, pp. 400–407.
- [10] N. Tax, N. Sidorova, R. Haakma, and W. M. P. van der Aalst, "Event abstraction for process mining using supervised learning techniques," in *Proceedings of SAI Intelligent Systems Conference (IntelliSys)*, ser. LNNS, vol. 15, Springer, 2016.
- [11] S. J. J. Leemans, K. Goel, and S. J. van Zelst, "Using multi-level information in hierarchical process mining: Balancing behavioural quality and model complexity," in *2nd International Conference on Process Mining, ICPM*, 2020, pp. 137–144.
- [12] F. Mannhardt, M. de Leoni, H. A. Reijers, W. M. P. van der Aalst, and P. J. Toussaint, "Guided process discovery - A pattern-based approach," *Information Systems*, vol. 76, pp. 1–18, 2018.
- [13] T. Baier, J. Mendling, and M. Weske, "Bridging abstraction layers in process mining," *Information Systems*, vol. 46, pp. 123–139, 2014.
- [14] G. Leonardi, M. Striani, S. Quaglini, A. Cavallini, and S. Montani, "Towards semantic process mining through knowledge-based trace abstraction," in *International Symposium on Data-Driven Process Discovery and Analysis*, ser. LNBIP, vol. 340, Springer, 2017, pp. 45–64.
- [15] R. Mans, W. M. P. van der Aalst, R. J. B. Vanwersch, and A. J. Moleman, "Process mining in healthcare: Data challenges when answering frequently posed questions," in *Process Support and Knowledge Representation in Health Care - BPM, Revised Selected Papers*, ser. LNCS, vol. 7738, Springer, 2012, pp. 140–153.
- [16] S. Remy, L. Pufahl, J.-P. Sachs, E. P. Böttinger, and M. Weske, "Event log generation in a health system: A case study," in *International Conference on Business Process Management*, ser. LNCS, Springer, 2020.
- [17] V. I. Levenshtein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals," *Soviet Physics Doklady*, vol. 10, p. 707, 1966.
- [18] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering block-structured process models from event logs containing infrequent behaviour," in *Business Process Management Workshops*, ser. LNBIP, vol. 171, Springer, 2013, pp. 66–78.