The Covert Ontology of Process Mining: Data-Driven Event Semantics

Riley Moher* and Michael Grüninger

Department of Mechanical and Industrial Engineering, University of Toronto,
Ontario, Canada M5S 3G8

Keywords: process mining · events · event log · process knowledge representation · ontology

*Corresponding author: Riley.Moher@mail.utoronto.ca

Process Mining: a New Perspective on Events

Process mining is a broad practice that transforms event logs into valuable measurements and aggregations of business process knowledge [1, 2, 3]. This includes tasks like process discovery (where a process model is inferred), concept drift (where changes in the way a process is executed over time is quantified), and predictive process monitoring (where the most likely next event within a process is estimated). These tasks require deep and diverse kinds of process knowledge to complete, yet they are all derived from the same kind of data, event logs. The eXtensible Event Stream (XES) Standard [4] is recognized as the standard format for defining event logs, and it defines an event as "an atomic granule of activity that has been observed". This standard also offers flexibility to further describe events with transitions like schedule, withdraw, suspend, or resume, and states like closed, running, obsolete, or reserved. We focus on the practice of process mining, however, as distinct from the standards or data of process mining; it is how process mining is actually done, not how it is said to be done or supposed to be done. This practice of process mining contains intended process/event semantics that are both novel and useful for process knowledge representation. We extend this thinking by establishing the following research claims:

1. The ontology of events, as defined in the practice of process mining, forms a useful and novel ontology of process knowledge.
2. Current event and process mining ontologies fail to capture the ontological commitments of the practice of process mining.
3. Real-world business process questions and associated data from the practice of process mining fulfill the role of guiding challenges in the development of practical process ontologies.

A Covert Event Ontology?

Consider the real-world example of the business process intelligence challenges (BPIC) [1], a series of real-world business challenge questions and accompanying event log datasets, widely used as benchmarks in process mining tasks [5, 6]. BPIC 2012 and 2017, for instance, pose challenges about the loan application process of a Dutch financial institute, including:

1. Identify which decisions have greater influence on the process flow.
2. What is the influence on the frequency of incompleteness to the final outcome?

Decisions, outcomes, process flow, correlation and causality: (at least) all of these concepts must be understood or defined to answer just these two questions. While the tools and methodologies used in solving business challenges with event log data may differ [6], they demonstrate shared ontological commitments (even if they are not made explicit). Some of these commitments may

[1] https://www.tf-pm.org/competitions-awards/bpi-challenge

be seen as "trivial", like events and timepoints being distinct classes, or events being countable (both of which are consistent with all BPIC analyses), though others are much less intuitive. Winning submissions of BPIC 2011 [7] and BPIC 2017 [8] both recognize that events in their logs can be subsumed by other events based on event-level attributes (patient diagnosis codes for BPIC 2011, and event name prefixes for 2017). Formalizing the semantics of these commitments allows these valuable business process insights to be more transparent, verifiable, and machine-readable.

Where Current Ontologies Differ

There are several ontologies proposed for use in process mining, including the SUPER Stack [1], EVO, and BPMO [9]. However, these applications are limited by reliance on domain ontologies, limited expressiveness, and scope limited by process modelling languages. Katsumi and Grüninger [10] also demonstrated expressiveness as an issue in three prominent event ontologies, Simple Event Model Ontology (SEM) [11], the ontology for Linking Open Descriptions of Events (LODE) [12], and the Event Ontology [13].

Upper ontologies that model events, including DOLCE [14] [15], UFO [16] and BFO [17] lack clear applications or extensions for the practice of process mining. Similarly, the Process Specification Language (PSL) [18], while having a more specific scope, lacks extensions to model diverse event kinds like aborted or suspended activities.

Design and Creation of Process Mining Event Ontologies

An ontology for the practice of process mining should provide a library of semantically-rich concepts to reason about events and the processes they describe. Benchmark problems to guide the development of such an ontology should be straightforward for practitioners to answer, while being non-trivial to represent in a formal, machine-readable way [19]. Furthermore, this ontology should be grounded in the data it is meant to represent and reason with. BPIC Challenge questions are an excellent pool to draw from as they require the understanding of complex process knowledge and drive many implicit ontological commitments. However, they are not specified as formal benchmark or challenge problems. To address this, we pose the formulation of challenge problems as:

1. $\mathcal{T}$ - a formal ontology of events and process mining concepts, relevant to the query
2. $\mathcal{T}_D$ - a domain ontology that is a Conservative extension of $\mathcal{T}$
3. $\mathcal{A}$ - a set of facts representing the event log(s) and domain process knowledge, expressed in the language of $\mathcal{T} \cup \mathcal{T}_D$

[1] https://web.archive.org/web/20091213192728/http://www.ip-super.org/component/option,com_frontpage/Itemid,1/

4. A query Q in the language of $\mathcal{T} \cup \mathcal{T}_D$, such that proving

$$\mathcal{T} \cup \mathcal{T}_D \cup \mathcal{A} \models Q$$

provides the suitable answer to the challenge problem.

Working through real-world business questions like BPIC will result in the development of a process ontology relevant for real process mining applications, with their common elements consisting of the essential concepts required to model event logs.

Conclusion and Future Work

We have motivated the development of novel event ontologies based on the covert ontology in the practice of process mining, and provided a framework to use guiding challenges in their development. In addition to a full specification of this methodology, the complete explanation will include axiomitizations of a modular ontology of events for the practice of process mining. This methodology of developing an ontology grounded in real-world datasets together with challenge problems and solutions, is also a framework that could provide value in the development in other domain ontologies.

References

[1] Wil Van Der Aalst et al. "Process mining manifesto". In: *Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9*. Springer. 2012, pp. 169–194.

[2] Wil Van Der Aalst. "Process mining: Overview and opportunities". In: *ACM Transactions on Management Information Systems (TMIS)* 3.2 (2012), pp. 1–17.

[3] Cleiton dos Santos Garcia et al. "Process mining techniques and applications–A systematic mapping study". In: *Expert Systems with Applications* 133 (2019), pp. 260–295.

[4] Christian W Gunther and HMW Verbeek. "Xes-standard definition". In: (2014).

[5] Malte Thiede, Daniel Fuerstenau, and Ana Paula Bezerra Barquet. "How is process mining technology used by organizations? A systematic literature review of empirical studies". In: *Business Process Management Journal* 24.4 (2018), pp. 900–922.

[6] Iezalde F Lopes and Diogo R Ferreira. "A survey of process mining competitions: the BPI challenges 2011–2018". In: *Business Process Management Workshops: BPM 2019 International Workshops, Vienna, Austria, September 1–6, 2019, Revised Selected Papers 17*. Springer. 2019, pp. 263–274.

[7] RP Jagadeesh Chandra Bose and Wil MP van der Aalst. "Analysis of Patient Treatment Procedures." In: *Business Process Management Workshops (1)*. Vol. 99. 2011, pp. 165–166.
[8] Elizaveta Povalyaeva, Ismail Khamitov, and Artyom Fomenko. "Bpic 2017: density analysis of the interaction with clients". In: *BPI Challenge* 2017 (2017).
[9] Carlos Pedrinaci and John Domingue. "Towards an ontology for process monitoring and mining". In: *CEUR Workshop Proceedings*. Vol. 251. 2007, pp. 76–87.
[10] Megan Katsumi and Michael Grüninger. "Using PSL to extend and evaluate event ontologies". In: *International Symposium on Rules and Rule Markup Languages for the Semantic Web*. Springer. 2015, pp. 225–240.
[11] Willem Robert Van Hage et al. "Design and use of the Simple Event Model (SEM)". In: *Journal of Web Semantics* 9.2 (2011), pp. 128–136.
[12] Ryan Shaw, Raphaël Troncy, and Lynda Hardman. "Lode: Linking open descriptions of events". In: *The Semantic Web: Fourth Asian Conference, ASWC 2009, Shanghai, China, December 6-9, 2009. Proceedings 4*. Springer. 2009, pp. 153–167.
[13] Yves Raimond and Mark Sandler. "Evaluation of the music ontology framework". In: *Extended Semantic Web Conference*. Springer. 2012, pp. 255–269.
[14] Antony Galin. "The Ontology of Time and Process". In: *Third Interdisciplinary School on Applied Ontology*. 2016.
[15] Stefano Borgo et al. "DOLCE: A descriptive ontology for linguistic and cognitive engineering". In: *Applied ontology* 17.1 (2022), pp. 45–69.
[16] Giancarlo Guizzardi et al. "UFO: Unified foundational ontology". In: *Applied ontology* 17.1 (2022), pp. 167–210.
[17] J Neil Otte, John Beverley, and Alan Ruttenberg. "BFO: Basic formal ontology". In: *Applied ontology* 17.1 (2022), pp. 17–43.
[18] Michael Grüninger. "Ontology of the process specification language". In: *Handbook on ontologies*. Springer, 2004, pp. 575–592.
[19] Ernest Davis. "How to write science questions that are easy for people and hard for computers". In: *AI magazine* 37.1 (2016), pp. 13–22.