

Analyzing Tools and Conceptual Models for Event Log Extraction from Relational Enterprise Database Systems

1 Introduction

Applications in process mining require standardized event-based data formats, called event logs. Examples for such standardized formats include XES (Group et al., 2016) or OCEL (Ghahfarokhi et al., 2020). However, companies typically manage their enterprise data in relational databases. Extracting event-based event logs from these databases necessitates meticulous data transformation. This is addressed in scientific literature with the introduction of numerous tools that aim to simplify the extraction of event logs from enterprise databases (Dakic et al., 2020). The problem of such tools is that they are often not maintained anymore, or they require follow-up work to be compatible with specific database systems.

Another line of work limits itself to the introduction of "metamodels" (See for example work by Pajić Simović et al. (2021)), which avoid specific implementations but provide practitioners with conceptual tools that simplify the identification and extraction of event-based data.

Whether metamodels or specific tools are provided, the corresponding literature explaining these typically involves an introduction of the underlying approaches on a conceptual level. Conceptual models are a common tool in software engineering to explain the desired properties and functionalities of software, independent of a specific implementation (Embley and Thalheim, 2012).

An initial investigation for this proposal revealed that tools previously developed by the scientific community for extracting event logs from enterprise databases are difficult to directly compare, as they are often not maintained to function in modern environments. In other words, many of these applications cannot be executed without extensive manual effort to debug these tools. However, such research can still be beneficial for practitioners since these tools are typically introduced with accompanying publications describing them on a conceptual level. Therefore, the goal of this work is to compare previous work on extracting event logs from enterprise databases on a conceptual level to generate insights on commonalities and differences between tools as well as focal points in the design of such tools. This enhances the comparability between presented approaches, allows to identify unaddressed research gaps for future work, and helps practitioners to generate own conceptual models for specific implementations.

2 Research Questions

The research questions are formulated as follows:

1. Which conceptual models are used in the scientific literature to describe the extraction of event logs from enterprise databases?
2. Are there different groups of work that highlight specific points but omit focal points of other groups in the conceptual description?
3. What do previously presented approaches have in common on a functional level, where do they differ, and where are current research gaps?
4. Can we recommend a conceptual model to which all of the previous work can be "translated" to in order to have one consistent representation that allows to compare the approaches on a functional level?

3 Work Packages

The specific work packages are:

- Complement a prior structured literature review by Dakic et al. (2020) to include not only specific implementations but also conceptual models (sometimes also called metamodels) and include work published in this area after the year 2019. The review shall be conducted based on the structured approach presented by Kitchenham (2004) including a full documentation.
- Analyze the collected work for conceptual models used and create a comprehensive overview.
- Discuss similarities and differences, strengths and weaknesses, as well as gaps/limitations of the different conceptual models. This is complemented by a research on conceptual modeling in software engineering to enhance the capability of comparing the models.
- Compare the found tools/metamodels on a functional level and discuss the capabilities of each tool to identify groups/clusters of approaches. The criteria/metrics for comparing the approaches will be defined based on the findings of the previous steps.
- Derive a proposed "best practice" to present conceptual models for extracting event logs from enterprise databases and describe all found tools/metamodels in terms of this best practice approach.
- Address the gaps of current approaches by building a conceptual model for a new tool which makes it easier for domain experts to extract high quality event logs.
- Optional: Compare and evaluate the existing models empirically against our proposed model and demonstrate its applicability on a case study with real business process data provided by Camelot ITLab GmbH. To perform this, we will derive SQL query patterns from the models, map them to an output format like XES as done by Pajić Simović et al. (2021) and analyse the quality of obtained event logs based on metrics like trustworthiness, accuracy, completeness, semantic clarity etc as defined by van der Aalst et al. (2012).

Advisors: Prof. Dr. Bjoern Eskofier
Johannes Roider
Dr. An Nguyen
Michael Simonis (Supervisor from Camelot ITLab)

Student: Sayantani Sensharma

Start – End: 1st May 2024 - 31st October 2024

References

- Dakic, D., Stefanovic, D., Lolic, T., Narandzic, D., and Simeunovic, N. (2020). Event log extraction for the purpose of process mining: a systematic literature review. In *Innovation in Sustainable Management and Entrepreneurship: 2019 International Symposium in Management (SIM2019)*, pages 299–312. Springer.
- Embley, D. W. and Thalheim, B. (2012). *Handbook of conceptual modeling: theory, practice, and research challenges*. Springer.
- Ghahfarokhi, A. F., Park, G., Berti, A., and van der Aalst, W. (2020). Ocel standard. *Process and Data Science Group, RWTH Aachen University, techreport*, 1.
- Group, X. W. et al. (2016). Ieee standard for extensible event stream (xes) for achieving interoperability in event logs and event streams. *IEEE Std*, 1849:1–50.

- Kitchenham, B. (2004). Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(2004):1–26.
- Pajić Simović, A., Babarogić, S., Pantelić, O., and Krstović, S. (2021). Towards a domain-specific modeling language for extracting event logs from erp systems. *Applied Sciences*, 11(12):5476.
- van der Aalst, W. et al. (2012). Process mining manifesto. In Daniel, F., Barkaoui, K., and Dustdar, S., editors, *Business Process Management Workshops*, pages 169–194, Berlin, Heidelberg. Springer Berlin Heidelberg.