

Interactive Process Identification and Selection from SAP ERP (Extended Abstract)

Julian Weber, Alessandro Berti^{1,2,*}, Gyunam Park¹, Majid Rafiei¹ and Wil van der Aalst^{1,2}

¹Process and Data Science Group @ RWTH Aachen, Aachen, Germany

²Fraunhofer Institute of Technology (FIT), Sankt Augustin, Germany

Abstract

SAP ERP is one of the most popular information systems supporting various organizational processes, e.g., O2C and P2P. However, the amount of processes and data contained in SAP ERP is enormous. Thus, the identification of the processes that are contained in a specific SAP instance, and the creation of a list of related tables is a significant challenge. Eventually, one needs to extract an event log for process mining purposes from SAP ERP. This demo paper shows the tool *Interactive SAP Explorer* that tackles the process identification and selection problem by encoding the relational structure of SAP ERP in a labeled property graph. Our approach allows asking complex process-related queries along with advanced representations of the relational structure.

Keywords

ETL, SAP, Object-Centric Process Mining

1. Introduction

Process mining is a branch of data science that provides methods for the analysis of event data recorded by information systems such as ERP and CRM systems. An essential step for such analyses is extracting an event log from the information systems. Moreover, this is one of the most time-consuming steps in most process mining projects. Thus, a successful extraction is a key to any process mining initiative.

The SAP ERP system is a popular choice for managing business processes such as order-to-cash (O2C, management of orders from the customers) and procure-to-pay (P2P, management of orders to the suppliers). Despite its widespread adoption, the extraction of event logs from an SAP ERP system remains ad-hoc. Commercial vendors such as Celonis provide extraction tools, focusing on the most common processes, e.g., O2C and P2P. However, other processes are under less attention, such as inventory management, financial planning, accounting, and production control processes, leading to few process mining projects in such processes.

It is challenging to extract event logs from SAP ERP systems due to their complexity, e.g., a typical SAP system contains 800,000 tables with tons of relationships. This inevitably requires

ICPM 2022 Doctoral Consortium and Tool Demonstration Track

*Corresponding author.

✉ julian.weber1@rwth-aachen.de (J. Weber); a.berti@pads.rwth-aachen.de (A. Berti);
gnpark@pads.rwth-aachen.de (G. Park); majid.rafiei@pads.rwth-aachen.de (M. Rafiei);
wvdaalst@pads.rwth-aachen.de (W. v. d. Aalst)



© 2022 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)

domain knowledge from the process experts of the organization. To this end, the expert needs to 1) identify the process to analyze, 2) select the relevant tables containing relevant data of the process, and 3) design query statements, e.g., using SQL.

In this demo paper, we present the tool *Interactive SAP Explorer* to support the domain expert for the first two initial steps, i.e., *process identification* and *tables selection*. Given a user input of a core document class in the organization, the tool identifies the most relevant process and the underlying tables. For instance, if the input by the user is *purchase order document*, then the most relevant process is the P2P process, and the underlying tables are as follows: *EBAN* for a purchase requisitions, *EKKO* for purchase orders, *EKBE* for goods/invoice receipts, *RBKP/RSEG* for invoice processing, *BKPF/BSEG* for payments, and *CDHDR/CDPOS* for changes in documents.

The tool first encodes the relational structure of SAP in a labeled property graph inserted inside a graph database. Then, a web interface is provided that permits the exploration of the relational structure of the SAP instance, the identification of the most important processes, and the creation of a list of tables for extraction. The list of tables is eventually provided to another component of the tool which has been previously introduced in [1] which creates an object-centric event log from such a list of tables. The tool improves the prototype proposed in [1] with better performance, customization, and exploration possibilities, particularly in the process identification and selection phases.

The rest of this extended abstract is organized as follows. Section 2 describes the functioning of the extractor. Section 3 points to the availability of the tool. Section 4 discusses the maturity of the tool. Eventually, Section 5 concludes the paper.

2. Innovations and Features

This section explains 1) *process identification and selection*, which is the main contribution of this paper, and 2) *process extraction*, which uses the output of this paper to produce event logs. First, the process identification and selection is implemented as follows:

- The elements of the relational structure of SAP that are important for the definition of a set of classes related to a given process are imported inside a graph database (Neo4J).
 - A graph database permits a faster exploration of the neighboring entities to a given concept because the connections are referenced inside the node object.
 - The chosen graph database (Neo4J) provides efficient implementations of layout algorithms, which can be executed on a significant amount of nodes/edges to provide an understandable graphical representation of the relational structure in SAP.
- Then, the identification process can be started. The first step is to identify a document type of interest (for example, the *purchase orders* and *sales orders*). This is directly connected, in the relational structure of SAP, to a set of tables (*purchase orders* are connected to the tables *EKKO*, *EKPO*, *EKPA*, *EKET*, *EKKN*).
- The next step is expanding the aforementioned set of tables. Starting from the initial set of tables, we identify the tables connected to the initial tables via the relational structure. The union of these tables contains the set of events regarding a process in SAP. For example, by expanding the tables related to the *purchase orders* document type, we get

a set of tables including purchase requisitions (*EBAN*), goods/invoice receipts (*EKBE*), accounting documents (*BKPF*), and other tables containing the events of the P2P process in SAP.

The process extraction component, which uses the approach described in [1], aims to extract an object-centric event log out of the SAP system based on the relevant tables identified in the previous step. There is no need to specify any SQL query.

- A pre-processing step is performed to restrict the extraction to the desired configuration.
- The extraction of the object-centric event log is performed, with an output following the OCEL specification <http://www.ocel-standard.org/>.

3. Availability of the Application

The source codes of the different components of the tool are available in the following repositories:

- *Layer of web services that can be run on IIS*: this component can be downloaded at <https://github.com/Javert899/interactive-extractor-from-sap-main/tree/main/Backend-C%23/SAPExtractorAPI>.
- *Angular web application*: this component can be downloaded at <https://github.com/Javert899/interactive-extractor-from-sap-main/tree/main/Frontend/InteractiveSAPExtractor>.
- *Python web services for the extraction of the object-centric event log*: this component can be downloaded at <https://github.com/Javert899/sap-extractor>.

Note that there is a dependency on non-open source UI components which need to be licensed to a single user. Therefore, the application is not directly runnable from the aforementioned source repositories. Also, the extractor requires the availability of an SAP ECC instance supported by the Oracle database and the installation of the Neo4J graph database, which is released under a proprietary license. The authors can provide access to the compiled version of the project under request. A videocast of the application is provided at the address <https://www.youtube.com/watch?v=Wi2xuUS0YSY>.

4. Maturity

The existing version of the tool can connect only to an SAP ECC instance supported by the Oracle database. Despite this being a popular option, this limits the possibility to apply the extractor in a generic setting. The extractor needs different components to run. This is architecturally complicated and, therefore, highly dependent on the functioning of existing queries/connectors on different versions of the software.

Our extractor overcomes the following limitations of existing SAP extractors; they are process-specific, they rely on traditional event logs, and suffer from convergence/divergence issues.

However, there are remaining limitations, including the fairly basic definition of activity/timestamp concepts. The choice of the graph database to navigate the relational structure of SAP is advantageous in terms of performance. After the selection of a set of tables, the extraction of an object-centric event log is left to the Python component, which executes many SQL queries to load the information needed in memory. Therefore, the extractor is limited by the amount of memory of the client.

The challenges are on both the theoretical and practical side. Theoretically, the selection of the activity concept is still a challenge. Practically, supporting different editions of SAP with different underlying databases, and an in-memory approach to compose the object-centric event log are still open challenges.

5. Conclusion

This demo paper presents an interactive extractor of object-centric event logs from SAP ERP, which is composed by two components: *process identification and selection* (novelty of this paper) and *process extraction* (using [1]). While the tool's code is open-source, it relies on some components released with a proprietary license. Section 4 discusses some limitations of the tool with the current architecture, which compromises its applicability in an enterprise setting

6. Acknowledgments

We thank the Alexander von Humboldt (AvH) Stiftung for supporting our research. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC-2023 Internet of Production – 390621612.

References

- [1] A. Berti, G. Park, M. Rafiei, W. M. P. van der Aalst, An event data extraction approach from SAP ERP for process mining, in: J. Munoz-Gama, X. Lu (Eds.), *Process Mining Workshops - ICPM 2021 International Workshops*, Eindhoven, The Netherlands, October 31 - November 4, 2021, Revised Selected Papers, volume 433 of *Lecture Notes in Business Information Processing*, Springer, 2021, pp. 255–267. doi:10.1007/978-3-030-98581-3_19.