Konekti: A Data Preparation Platform for Process Mining (Extended Abstract)

Lotte Vugs[1], Maarten van Asseldonk[1] and Niek van Son[1]

[1]*Waves Process Intelligence, Eindhoven, The Netherlands*

Abstract

Data preprocessing frequently consumes a large proportion of the time spent on process mining projects. Although several commercial process mining software solutions offer data preparation scripts, these typically cover only a narrow set of process perspectives and source systems. When a particular perspective or system is not supported, practitioners usually need to fall back on ad-hoc ETL solutions. This paper introduces Konekti: a tool dedicated to offering a structured approach to prepare event data for process mining.

Keywords

Process Mining, Data Preparation, Data Transformation, Object-Centric, System-agnostic

1. Introduction

Process mining consists of a set of methods, tools and techniques to discover process models, analyze performance, check compliance and compare variants using event data recorded in information systems [1]. The key input for process mining is an *event log*, a collection of events corresponding to a specific process. In practice, event logs are usually not readily available. Instead, event data needs to be selected and extracted from a set of tables (or systems), and then cleaned and transformed in order to create an event log suitable for analysis. Depending on the complexity of the source data model and process to be discovered, up to 80% of a project time span is used for data preprocessing and log creation, leaving 20% for the process analytical work [2].

Thus far, most scientific research has focused on data analysis rather than data preprocessing [2]. The preprocessing tools that do stem from scientific research are usually prototypes and do not seem to be used widely by practitioners: none of them were mentioned as tools used for data pre-processing in the recent XES survey [3], nor were they mentioned in the interviews with practitioners we held as part of a user tests. At the same time, commercial process mining software primarily offers *connectors*: specific procedures that define for a particular source system and process perspective how to extract relevant data, and which additional transformations should be applied [4]. Practitioners who wish to preprocess event data from a source and/or process for which their software vendor has no *connector* usually need to fall back on ad-hoc solutions, like SQL editor tools.

This paper introduces Konekti, a no-code tool for preprocessing event data for process mining. Konekti supports practitioners in preprocessing their process data. Its output, an event log, can be directly used as input for existing process mining tools, making it an ideal tool to be used in tandem with existing solutions.

2. Related Work

Several academic tools focus on extracting event data from operational systems and transforming it into event logs. These include the EVS Model Builder [5], Xtract [6], and Eventifier [7]. An underlying assumption of existing tools is that there is a single notion of a *case*. However, processes in ERP systems are usually organized around multiple objects having one-to-many and many-to-many relations. Often, more than one object could potentially serve as a case notion. Which object is the best candidate depends on the perspective of interest, which can change over time.

Recently, more research has been done on how to extract and transform *object-centric* data in a way that offers more flexibility in changing the case notion. A noteworthy contribution in this field is the Onprom tool [8]. Furthermore, the Object-Centric Event Data standard [9] is currently being developed. Although designed independently, Konekti's meta-model mimics the proposed standard almost completely, apart from some semantic details.

3. Tool Outline

Konekti is a tool dedicated to offer a structured approach to prepare event data. Konekti offers some key benefits over the tools currently in use by practitioners:

- **System-agnostic**: commercial process mining tools focus on data preparation scripts for a specific process perspective in a specific source system. Konekti focuses on simplifying data preparation in case the available scripts do not match with the process perspective and/or source systems in scope.
- **Built for practitioners**; in contrast to academic prototypes, Konekti is designed to be used in industrial settings. Therefore, in developing the platform, a strong emphasis is put on user-friendliness.
- **No-code**: in contrast to SQL editor tools, and most tools stemming from academic research, Konekti is completely no-code. This removes the hurdle of learning a new coding language, and stimulates shareability across users with different levels of coding skills.
- **Object-centric**: rather than assuming the case notion is fixed at the start of the data preparation, Konekti implements an object-centric data model. From this data model, different event logs can be exported that use different case notions.

A screencast demonstrating the tool is available online[1].

---

[1] https://youtu.be/fjjexHpl9KM

Main Components A *connector* is the link to a specific database schema from which a user can extract event data. A user can set up a connector by specifying the connection details in a form.

In the *event store*, the user identifies which data needs to be extracted and, simultaneously, maps this data to Konekti's object-centric data model. Konekti's meta-model is built around two main artifacts. The first of them is the *object*, representing a type of information carrier in a process. The second main artifact is the *activity*, representing a mutation of an object. Objects can be related to other objects using object-object relations. Similarly, objects can be related to activities using object-activity relations.

Configuration of the artifacts in the event store is done using forms rather than scripts (i.e. *no-code*). After configuration, a query is automatically generated and a materialized view of the data is stored in Konekti's database.

In the *export builder*, the user configures event logs from the data in the event store. Depending on the perspective of interest, the user selects which object is used as case notion. In the background, a breadth-first search is executed over all object-object relations to generate a list of related objects. From the list of related objects, the user selects the objects to include. Then, the user selects the activities to include; here the user can select activities that have an object-activity relation to the case object or to any of the related objects selected previously. Third, the user can select which attributes to include. *Event* attributes can be selected from activity attributes. *Case* attributes can be selected from the case object. After configuration, a single event log table is materialized and can be downloaded as a CSV file.

Limitations Konekti will be released in the beginning of 2023; the version described in this paper is a prototype under active development. In this version, data can only be extracted from PostgreSQL and MSSQL databases, making it more difficult to apply Konekti to data stored in other databases. Also, the process of setting up an event store still contains some unnecessarily repetitive tasks. Furthermore, there are no case studies done with Konekti yet. Test databases from several ERP systems have been used to test the tool, but it is not yet implemented in a practical setting. Additionally, the features to transform the data during preprocessing are still limited. To work around this, users can pre- or post-process the event data before/after loading it in Konekti.

4. Future Work

We are planning to launch Konekti in the beginning of 2023. Until the launch, we will focus on implementing Konekti in practice, and expanding its features (e.g., extending the connector types and export formats and including activity aggregation). Furthermore, we're planning to augment and improve the platform based on the feedback gathered from this conference and alignment with the OCED working group.

Once the foundations for event data preprocessing are laid, we plan to focus on enabling and promoting the exchange of data preprocessing knowledge within the process mining community. To enable that, all Konekti users agree to share the steps taken in the data preprocessing will be eligible for a free license.

5. Conclusion

The event log is the fundamental input for every process mining analysis, yet creating one can be difficult and time consuming. Practitioners who wish to analyze a process from a source system that is not supported by their process mining tool of choice are often forced to fall back on ad-hoc solutions. This paper introduces Konekti: a user-friendly, no-code platform that supports practitioners in data preprocessing, regardless of the systems included in the scope of their analysis.

Acknowledgments

References

[1] W. Van der Aalst, Process mining: Data science in action, 2016. doi:10.1007/978-3-662-49851-4.

[2] R. Accorsi, J. Lebherz, A practitioner's view on process mining adoption, event log engineering and data challenges, in: Process Mining Handbook, Springer, Cham, 2022, pp. 212–240.

[3] M. T. Wynn, J. Lebherz, W. M. van der Aalst, R. Accorsi, C. Di Ciccio, L. Jayarathna, H. Verbeek, Rethinking the input for process mining: insights from the xes survey and workshop, in: International Conference on Process Mining, Springer, 2022, pp. 3–16.

[4] J. De Weerdt, M. T. Wynn, Foundations of Process Event Data, Springer International Publishing, Cham, 2022, pp. 193–211. URL: https://doi.org/10.1007/978-3-031-08848-3_6. doi:10.1007/978-3-031-08848-3_6.

[5] J. E. Ingvaldsen, J. A. Gulla, Preprocessing support for large scale process mining of sap transactions, in: International Conference on Business process management, Springer, 2007, pp. 30–41.

[6] E. H. Nooijen, B. F. v. Dongen, D. Fahland, Automatic discovery of data-centric and artifact-centric processes, in: International conference on business process management, Springer, 2012, pp. 316–327.

[7] C. Rodrıguez, R. Engel, G. Kostoska, F. Daniel, F. Casati, M. Aimar, et al., Eventifier: Extracting process execution logs from operational databases, Proceedings of the demonstration track of BPM 940 (2012) 17–22.

[8] D. Calvanese, T. E. Kalayci, M. Montali, S. Tinella, Ontology-based data access for extracting event logs from legacy data: the onprom tool and methodology, in: International Conference on Business Information Systems, Springer, 2017, pp. 220–236.

[9] I. T. F. o. P. M. OCED working group, Object-centric event data: A meta model, Internal Document (2022).