

PROCESSEXPLORER: Interactive Visual Exploration of Event Logs with Analysis Guidance

Alexander Seeliger, Maximilian Ratzke, Timo Nolle, Max Mühlhäuser
Technische Universität Darmstadt
Telecooperation Lab
Darmstadt, Germany
Email: {seeliger, nolle, max}@tk.tu-darmstadt.de

Abstract—Process analysts use process mining techniques to obtain fact-based knowledge from event logs about how business processes are actually executed in organizations. Often process discovery is the first step in their analytical workflow. However, when working with large amount of data and complex processes, exploring as-is process models to obtain interesting and insightful knowledge can be challenging. We propose PROCESSEXPLORER, an interactive visual recommendation system for process discovery to facilitate event log exploration. PROCESSEXPLORER automatically analyzes the event log to obtain promising subsets of cases, evaluates interesting process performance indicators, and recommends those that are most interesting and insightful. Our system uses multi-perspective trace clustering to identify candidate cases of interest and a deviation-based approach to assess the interestingness of process performance indicators. We implemented PROCESSEXPLORER as a standalone desktop application that allows to explore any process and any event log. Our demo shows how the workflow of analysts is supported by the system through suggesting subset and insights recommendations.

Index Terms—process discovery, variants analysis, log pre-processing, trace clustering, statistical hypothesis testing

I. INTRODUCTION

Nowadays, information systems in organizations support and automate the processing of business transactions. These systems are typically integrated into companies' business processes and record the activities that have been executed in the form of an event log. Process mining aims at providing an accurate view of how processes are actually executed in organizations. In particular, process discovery reconstructs as-is process models from event logs which can be used for further analysis. A wide range of process mining tools has been established that implement process discovery and analysis methods to support analysts to obtain valuable knowledge. With this knowledge, process issues can be identified and optimizations can be implemented.

In this paper, we introduce the PROCESSEXPLORER system which provides recommendations to the analyst on how to select a subset of cases and what statistics may be interesting and insightful. Our system is inspired by the workflow that analysts typically perform when working with process mining tools. The visual inspection of the discovered process model is the initial starting point of any process mining project. Due to the massive growth of data, the increasing process complexity, and the flexible execution of business processes in

organizations, visual exploration and analysis are getting more and more challenging. Often the analyst is confronted with a spaghetti-like process map which by itself does not necessarily lead to useful insights. Without extensive knowledge about the underlying process, selecting the right set of cases to find interesting and valuable insights or trends is non-trivial. In current process mining tools, most of these analysis steps are performed manually, leading to a lot of repetitive work which hampers efficient exploration and analysis.

PROCESSEXPLORER extends the interactive visual exploration capabilities in today's process mining tools by providing automatic guidance to the analyst. Our tool integrates several recommendation suggestions in a user-friendly manner to improve overall process discovery exploration:

- 1) *Subset Recommendation.* PROCESSEXPLORER recommends subsets of interesting cases to allow analysts quickly inspect the different process behaviors observed in the event log. Different from the manual filtering that requires expert knowledge, subset recommendations are automatically derived by mining process behavior patterns from the dataset to simplify subset selection.
- 2) *Insights Recommendation.* After selecting a subset of cases, PROCESSEXPLORER automatically computes a range of relevant process performance indicators to show interesting deviations. Analysts are guided towards interesting statistics that they usually would compute manually.
- 3) *Recommendation Ranking.* In order to prevent the analyst from inspecting only a limited subset of cases, PROCESSEXPLORER provides the analyst with the most diversifying recommendations by applying diversifying top-k ranking [1].

PROCESSEXPLORER is agnostic to the process and event log that is being analyzed. Any process and any event log in the standardized IEEE XES format can be used. Furthermore, the analyst does not need to setup any configuration or specify parameter values. Prior knowledge about the process or the event log is not required. PROCESSEXPLORER obtains all the necessary information from the event log itself.

We used PROCESSEXPLORER in a case study on the BPI Challenge 2019 event log collected from a large company to investigate the procurement handling process [2]. The

rest of the paper is structured as follows. We provide a walk-through of PROCESSEXPLORER using this event log, showing the different types of recommendations provided by PROCESSEXPLORER and highlight the maturity of the tool. Then, we present the architecture of PROCESSEXPLORER to show extensibility.

II. RECOMMENDATION ENGINE

PROCESSEXPLORER extends process mining tools by introducing a recommendation engine to support analysts selecting interesting subsets of cases and generating insightful statistics. In particular, our system allows to quickly scan unknown processes in event logs to obtain knowledge about how the process is actually executed and where potential issues can be found. PROCESSEXPLORER provides two types of recommendations and a ranking mechanism.

A. Subset Recommendations

The first type of recommendation suggests subsets of cases that contain interesting process behavior patterns. We are particularly interested in patterns that combine the control flow and the data perspective. This is inspired by the manual work of analysts who not only filter cases by the sequence of activities but also by attributes. This is often used to compare different departments, products, or company locations. To support analysts during the selection of appropriate subsets of cases, PROCESSEXPLORER automatically analyzes the given event log to find such patterns using trace clustering. Specifically, we apply multi-perspective trace clustering [3] to obtain subsets of cases that contain dependencies between the control flow and the case attributes. Resulting subsets of cases with similar behavior lead to process maps that are typically less complex and easier to understand visually.

B. Insights Recommendations

Another typical task in process mining is to investigate and compare a range of process performance indicators (PPIs), such as the number of activities, the total duration time, the duration time between activities, the directly followed-by relation, and the existence of activities. These are either directly visualized in the process map or separately displayed in the form of statistical charts or single values. Existing process mining tools provide assistance by offering the possibility to create dashboards with predefined PPIs which will update immediately if a different case selection is made. Still, each PPI needs to be investigated one after another to identify deviations which is time-consuming and error-prone. PROCESSEXPLORER automatically computes these PPIs for a selected subset and identifies those ones that may be interesting to the user by performing statistical significance testing. Compared to dashboards that are static with respect to the computed PPIs, PROCESSEXPLORER reevaluates the PPIs for each applied subset recommendation individually. Only PPIs that are significantly different from the rest of the cases in the event log are considered as an interesting insight [4].

C. Ranking

Lastly, PROCESSEXPLORER ranks the recommendations based on the interestingness score [4]. Each insights recommendation is assigned a score that is computed from how large the deviation is from the rest of the event log and the number of cases that are covered. We use Cohen's effect size [5] which uses a comprehensive scale to determine the maturity of the deviation. Insights recommendations are then ranked by their assigned scores.

During our experiments, we found that certain insights co-occur with each other which unnecessarily increases the number of insights recommendations. PROCESSEXPLORER clusters similar insights recommendations using the Spearman's rank-order correlation.

Subset recommendations are assigned a score based on the insights scores and the number of cases that are contained in the subset. We obtain the top-k subset recommendations using the top-k diversifying ranking algorithm [1] to increase the analysts perspective on the event log. Instead of showing very similar subset recommendations on top of the list, PROCESSEXPLORER suggests the most diversifying subsets which prevent the analyst from inspecting only a limited subset of cases. In PROCESSEXPLORER, the top 10 most interesting and diversifying subset recommendations are shown to the user.

III. TOOL

PROCESSEXPLORER is a standalone interactive process mining tool to demonstrate the proposed guidance capabilities. As mentioned earlier, it allows importing any standardized IEEE XES event log and works without specifying any additional parameter value. We give a walk-through of PROCESSEXPLORER by inspecting the procurement handling process of the BPI Challenge 2019 event log [2]. Figure 1 shows the main screen of PROCESSEXPLORER. The user interface consists of five different components:

a) Process Map: The most prominent component in PROCESSEXPLORER is the process map. It visualizes the activities and transitions that have been observed in the event log. Activities and transitions can be filtered by their relative occurrence using the slider at the bottom right. Figure 1 shows the process map of a selected subset recommendation.

b) Subset Recommendations: On the top right side, the ranked list of subset recommendations is shown. Subset recommendations can be modified and adjusted by the user, enabling to further refine the selection of cases interactively. Users can add a happy path filter, a variant filter, a start and end activity filter, and an activity occurrence filter. Figure 1 shows the 8 subset recommendations that are suggested for the currently selected subset of cases.

c) Subset Statistics: On the lower right side, basic statistics of the selected subset recommendation are shown which give an overview of the cases in the subset. The statistics show how the subset selection compares to the original event log and highlights the event distribution, the variant distribution, and the number of selected cases. Based on the statistics, the

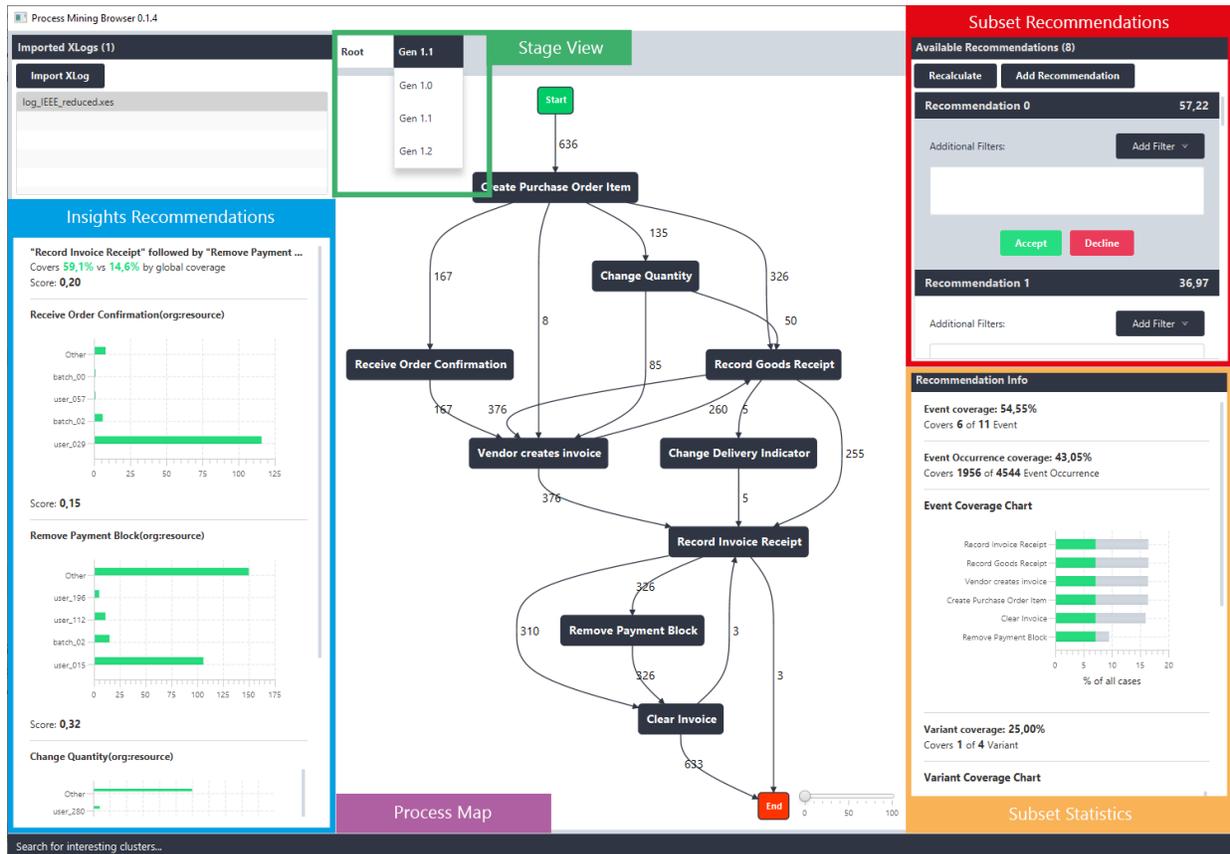


Fig. 1. User interface of PROCESSEXPLORER showing the subset and insights recommendations, the process map of the selected subset, the stage view, and the subset statistics. The screenshot shows a selected subset recommendation of the BPI Challenge 2019 event log.

user can decide which subset recommendation to apply. In the example, the selected subset recommendation selects 6 events, and 1 out of 4 variants.

d) *Insights Recommendations*: On the left-hand side, PROCESSEXPLORER shows the insights recommendations for the current subset. Insights recommendations are automatically updated each time the subset of cases is modified. The system computes a range of basic PPIs which are typically analyzed by users. We distinguish between case- and subset-based insights. Depending on the insight type, a different visualization is shown to the user. Figure 1 shows a portion of the obtained insights recommendations. For instance, the first insight refers to the directly followed-by relation between the “Record Invoice Receipt” and “Remove Payment Block” activities, which occurs more often in the applied subset. Furthermore, we can see that the activity “Receive Order Confirmation” is mostly executed by “user_029”.

e) *Stage Views*: For easier navigation between the different subset recommendations, PROCESSEXPLORER introduces stage views. Each time the user decides to apply a subset recommendation a new stage view is generated. A stage view stores the selected cases and the computed insight recommendations. Stages are organized as a hierarchical structure such that each refinement of a selection results in a new hierarchy level. For each stage view, subset and insights

recommendations are computed, so recommendations can be successively refined.

IV. ARCHITECTURE

PROCESSEXPLORER is built of three main components: the event log manager (*XLogManager*), the stage manager (*XStageManager*), and the recommendation manager (*RecommendationManager*). All three components are open for extension, such that other event log formats, stage management capabilities, subset and insights recommendation approaches can be integrated. Figure 2 shows the overall architecture of PROCESSEXPLORER.

Event logs are imported as an OpenXES *XLog* object and stored in-memory using the XESlite extension. Each loaded log is stored in the *XLogData* object structure which links to the *XLog* object and stores the basic statistics of the log. The *XStageManager* is responsible for managing the views of PROCESSEXPLORER, storing a history of all stages visited by the user. For an active stage, the *XStageManager* retrieves the recommendations from the *RecommendationManager* which returns a set of *Recommendation* objects. If the recommendations have not yet being computed, the *RecommendationManager* calls the *RecommendationFactory*. Each *Recommendation* refers to the subset recommendations

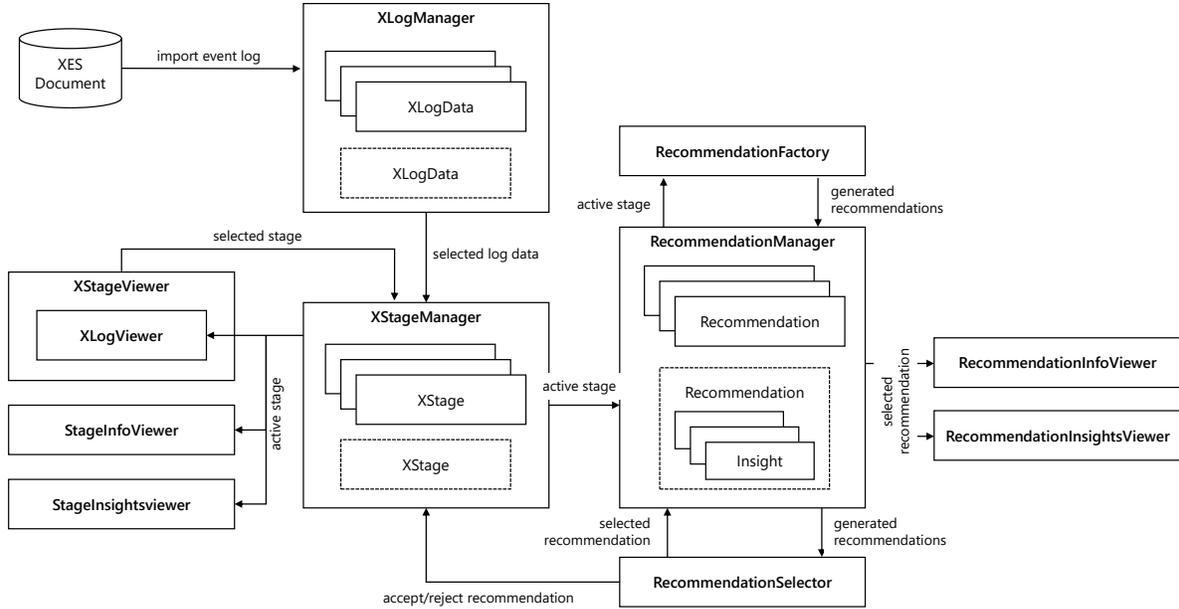


Fig. 2. Overview of the architecture of the PROCESSEXPLORER tool [6].

shown in PROCESSEXPLORER which contain the *Insight* recommendations.

All visualization components, such as the *XLogViewer*, *StageInfoViewer*, *StageInsightsViewer*, *RecommendationViewers* are separated from the actual recommendation engine. This architecture allows the exploration of different types of visualizations, such as other types of charts, process model visualizations, etc., but keep the actual computation of the recommendations.

In the current implementation of PROCESSEXPLORER, we implemented a multi-perspective trace clustering recommendation engine for subset recommendations and a statistical significance testing approach for obtaining insights recommendations. However, other implementations are easy to implement by extending the corresponding classes.

V. DOWNLOAD, SCREENCAST, AND LINKS

The PROCESSEXPLORER demo tool can be found at our project page¹. On the project page, a demonstration video including a screencast, a reduced event log derived from the BPI Challenge 2019, and additional screenshots are provided. The demo tool requires Oracle Java 8 and was tested on Windows and Ubuntu.

VI. CONCLUSION

In this paper, we presented PROCESSEXPLORER, an interactive visual recommendation system for process discovery inspired by the workflow typically performed by analysts. Our system suggests two types of recommendations that guide

analysts towards interesting subsets of cases as well as shows insightful statistics of relevant PPIs. Subset recommendations are computed using multi-perspective trace clustering to obtain process behavior patterns that are interesting to explore. Insights recommendations show interesting PPIs that significantly differ for an investigated subset compared to the rest of the event log. Furthermore, PROCESSEXPLORER gives each recommendation a score based on interestingness and maturity. It applies top-k diversifying ranking to obtain the most different recommendations.

ACKNOWLEDGMENT

This work is funded by the German Federal Ministry of Education and Research (BMBF) Software Campus project “AI-PM” [01IS17050] and the research project “KI.RPA” [01IS18022D].

REFERENCES

- [1] L. Qin, J. X. Yu, and L. Chang, “Diversifying top-k results,” *Proceedings of the VLDB Endowment*, vol. 5, no. 11, pp. 1124–1135, jul 2012.
- [2] B. F. van Dongen, “Dataset BPI Challenge 2019,” *4TU.Centre for Research Data*, 2019.
- [3] A. Seeliger, T. Nolle, and M. Mühlhäuser, “Finding Structure in the Unstructured: Hybrid Feature Set Clustering for Process Discovery,” in *Proc. of the 16th BPM*. Springer International Publishing, 2018, pp. 288–304.
- [4] M. Vartak, S. Rahman, S. Madden, A. Parameswaran, and N. Polyzotis, “SeeDB,” in *Proc. of the VLDB Endowment*, vol. 8, no. 13, 2015, pp. 2182–2193.
- [5] J. Cohen, “Statistical Power Analysis,” *Current Directions in Psychological Science*, vol. 1, no. 3, pp. 98–101, jun 1992.
- [6] M. Ratzke, “Intelligent and Systematic Browsing through Process Mining Data,” 2019.

¹<https://fileserv.tu.informatik.tu-darmstadt.de/AS/processexplorer>