Topic 1: Object-Centric Process Mining

The goal of process mining is to gain insights from event logs that record real-life process executions. In traditional process mining, these event logs have a single case notion that determines what a process instance is. For example, if the case notion is a sales order, then all events related to "Order 1" are assumed to belong to the same instance of a process. However, in real-life processes, events typically relate to several entities at the same time – for instance, an order may include multiple packages and items and be associated with a customer. "Flattening" this interplay of entities by assigning events to a single case notion provides only a limited view on the real process and can lead to convergence and divergence issues in the resulting event log. Object-centric process mining aims to address this by getting rid of the single case notion and developing new event log formats and process mining techniques.

The objective of this seminar thesis is to examine the current state of the art of object-centric process mining through a structured literature review. This involves identifying, for example, object-centric data formats and algorithms for process discovery or conformance checking and analyzing how they differ from traditional ones. Based on the results of the review, the participant should then pinpoint research gaps, i.e., determine for which aspects of process mining there are currently no object-centric counterparts or alternatives.

Literature:
- van der Aalst, W.M.P. (2019). Object-Centric Process Mining: Dealing with Divergence and Convergence in Event Data. In: Software Engineering and Formal Methods. Springer. 3-25.
- Ghahfarokhi, A.F., Park, G., Berti, A., & van der Aalst, W.M.P. (2021). OCEL: A Standard for Object-Centric Event Logs. In: New Trends in Database and Information Systems. Springer. 169-175.

Topic 2: Ex-Post-Analysis of Conformance Checking Results

Conformance checking is one of the main tasks of process mining. It comprises techniques for checking the relation between a designed process model and the real-life behavior of a process, with the goal of identifying and analyzing deviations between them. State-of-the-art conformance checking approaches have become computationally very efficient and robust to many kinds of process behavior. However, as the conformance checking itself cannot be considered as a means to an end, further analyses can prove useful in order to explain why deviations from the designed process model occur. Users will be enabled to improve the process conformance by that.

The goal of this seminar thesis is to identify approaches that analyze differences in conformance. Therefore, the participant should conduct a structured literature review regarding the ex-post usage of conformance checking results. The applicant should identify 1) which level of granularity in conformance measures is used (i.e., fitness, individual deviations, or deviation patterns like swapped activities), 2) for what purpose the conformance is analyzed, and 3) which methods are used to explain conformance differences. The review should be conducted wholistically, and the participant should contact the supervisor in case there is too much relevant literature (in which case the scope will be redefined). In a second step, the identified approaches should be analyzed with regard to a research gap or patterns in usage of conformance checking results.

Literature:
- M. de Leoni, M., van der Aalst, W.M.P., & Dees, M. (2016). A general process mining framework for correlating, predicting and clustering dynamic behavior based on event logs. Information Systems 56. 235-257.
- Accorsi, R., & Stocker, T. (2012). On the exploitation of process mining for security audits: the conformance checking case. In: Proceedings of the 27th Annual ACM Symposium on Applied Computing. Association for Computing Machinery. 1709–1716.

Topic 3: State-of-the-Art in Conformance Checking Techniques

Conformance checking comprises an important domain in process mining. Several techniques are widely known (e.g., rule checking, token-replay, alignments). Typically, they focus on the control-flow (i.e., the correct ordering of activities). Certain extensions have been developed to for example incorporate the data perspective into the conformance check.

The goal of this seminar thesis is to identify techniques that conduct the conformance check. A technique thereby can be alignments but also the extension of data-aware alignments. Therefore, the participant should conduct a structured literature review. The applicant should identify 1) what are the inputs and outputs of the technique, 2) what are requirements or limitations of each technique, and 3) what are typical applications of each technique. Thereby, also the extensions of each technique are of interest. The review should be conducted wholistically, and the participant should contact the supervisor in case there is too much relevant literature (in which case the scope will be redefined). In a second step, the applicant should compare the found techniques and elaborate on (dis-) advantages. Further, possible combinations of the techniques should be discussed.

Literature:
- Dunzer, S., Stierle, M., Matzner, M., & Baier, S. (2019). Conformance checking: a state-of-the-art literature review. In: Proceedings of the 11th International Conference on Subject-Oriented Business Process Management. Association for Computing Machinery. 1–10.
- Burattin, A., & Carmona, J. (2018). A Framework for Online Conformance Checking. In: Business Process Management Workshops (BPM 2017). Springer.

Topic 4: Context-Aware Business Process Prediction

Process mining is a research field that aims to provide techniques for generating useful insights into how processes are conducted from the event log data of process executions recorded by information systems. An emerging trend in this field is predictive process monitoring which is concerned with the prediction of the future behavior of a running process based on the analysis of its current behavior. Many of the predictive process monitoring techniques do not consider the fact that processes are executed within a particular context that affects their behavior, and this context is not usually considered during process analysis and prediction. Examples of a process context could be the amount of the invoice handled, the priority of a customer, or the role of a resource. Including process context in the analysis will lead to more informed insights about processes especially when it comes to predicting and recommending the remaining execution of a running process. For this reason, recent work in predictive process monitoring started to incorporate process context in process predictions.

The goal of this seminar thesis is to systematically review the current approaches for incorporating process context in process predictions. Accordingly, the participant is expected to conduct a structured literature review (SLR) to identify predictive process monitoring approaches incorporating process context in predictions. The review should cover 1) what type(s) of context is incorporated in the prediction, 2) the target of the prediction (e.g., predicting process outcome, or next activity, etc.), 3) details on how the context is incorporated (e.g., what type of data needed, techniques used, etc.), and 4) the impact of incorporated context on prediction results.

Literature:
- Brunk, J. (2020). Structuring business process context information for process monitoring and prediction. In: IEEE 22nd Conference on Business Informatics. IEEE. 39-48.
- Márquez-Chamorro, A. E., Revoredo, K., Resinas, M., Del-Rio-Ortega, A., Santoro, F. M., & Ruiz-Cortés, A. (2020). Context-aware process performance indicator prediction. IEEE Access, 8, 222050-222063.

Topic 5: The Influence of Sampling on Process Mining Results

Process mining is a family of techniques for the analysis of event logs that record the behavior of business processes, as captured from enterprise information systems. Especially for large systems, such as ERP systems, and frequently executed processes, such as procurement processes, these event logs can become very large. Some process mining techniques, especially computationally expensive ones like alignment-based conformance checking, are not able to handle these large event logs. To address this issue, some researchers use event log sampling: They select a small subset of traces from the large event log and use this subset as a proxy for the original log. Due to the smaller size, even more computationally expensive process mining techniques can produce meaningful results for these logs.

However, sampling from an event log might change important properties of the log, such as the contained activities, the number of variants, or their frequency distribution. These effects are currently not sufficiently assessed by process mining researchers. The goal of this seminar thesis is therefore to systematically explore (1) the sampling techniques that are currently employed in process mining research and (2) the detrimental effects that those techniques have on the results of applied process mining techniques. The participant should conduct two structured literature reviews. The first one is meant to identify the sampling techniques, whereas the second one is meant to find papers that have studied potential detrimental effects. Eventually, the thesis should make recommendations on which sampling technique is better suited for applications in different use cases and which effects of sampling should be empirically studied in the future.

Literature:
- Bauer, M., van der Aa, H., & Weidlich, M. (2022). Sampling and approximation techniques for efficient process conformance checking. Information Systems 104, 101666.
- Fani Sani, M., van Zelst, S.J., & van der Aalst, W.M.P. (2021). The impact of biased sampling of event logs on the performance of process discovery. Computing 103, 1085–1104.
- Knols, B., & van der Werf, J.M.E.M. (2019). Measuring the Behavioral Quality of Log Sampling. International Conference on Process Mining (ICPM) 2019. IEEE. 97-104.

Topic 6: Evaluating the Results of Process Discovery

Process discovery is perhaps the most commonly used process mining technique. Its purpose is to derive a process model from an event log that analysts can then use to understand and improve the execution of processes in an organization. Various process discovery algorithms exist, ranging from the simple alpha miner to the more powerful heuristic and inductive miners. While there are generally agreed upon quality criteria for discovered process models (soundness, fitness, precision, simplicity, and generalization), the way these criteria are measured is far less uniform.

In this seminar thesis, the participant should review process discovery literature in order to identify the various ways that discovery algorithms can be evaluated. The results should be categorized along the five quality criteria and be analyzed with regard to the general approaches (e.g., replay-based / alignments, statistics) and concrete measures employed.

Literature:
- Buijs, J.C.A.M., van Dongen, B.F., & van der Aalst, W.M.P. (2012). On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery. In: On the Move to Meaningful Internet Systems: OTM 2012. Springer. 305-322.
- Janssenswillen, G., Donders, N., Jouck, T., & Depaire, B. (2017). A comparative study of existing quality measures for process discovery. Information Systems 71, 1-15.

Topic 7: The Application of Process Mining to Non-Process-Related Data

Process mining focuses on the analysis of process event logs, i.e., sequences of event-based, timestamped datapoints that refer to the execution of a business process in an information system. However, from a purely data-centric perspective, process mining can also be applied to event logs that do not refer to the execution of a business process, but to other event-based longitudinal phenomena. Examples for these phenomena include medical trials, user interactions with a graphical user interface, activities of an autonomous robot, or even the movement of tectonic plates.

The goal of this seminar thesis is to provide a systematic overview of the application of process mining in non-process-related settings. Therefore, the participant should conduct a structured literature review to identify all recent papers that apply process mining to study phenomena that are not business processes in a traditional sense. Those papers should be analyzed regarding (1) the phenomenon in focus, (2) the data they require, and (3) the process mining techniques they apply.

Literature:
- Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(1), e1230.
- dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., ... & Scalabrin, E. E. (2019). Process mining techniques and applications–A systematic mapping study. Expert Systems with Applications, 133, 260-295.
- Emamjome, F., Andrews, R., & ter Hofstede, A. H. (2019). A case study lens on process mining in practice. In: On the Move to Meaningful Internet Systems: OTM 2019. Springer. 127-145.

Topic 8: Statistical Analysis in Process Mining

Process mining provides a set of techniques for the analysis of business process behavior as captured in the event log data of process executions recorded by information systems. Those techniques focus mainly on representing process behavior in a form of a process model covering the control-flow perspective of the process or an enhanced process model covering other aspects of the process such as the resource and time perspectives. Although representing process behavior over process models is important, having quantifiable measures representing different patterns in process executions can help managers have more informed insights about the process behavior and can even help in predicting the behavior of a running process based on different quantifiable metrics. Recent work in process mining started to incorporate different statistical techniques, such as regression, to provide more quantifiable process insights than just log statistics, especially in the area of predictive process monitoring where the future behavior of a running process is predicted based on the analysis of its current behavior.

The goal of this seminar thesis is to systematically identify and review the current state of the art regarding the application of regression as a statistical analysis technique in process mining. Therefore, the participant is expected to conduct a structured literature review (SLR) to identify all recent papers that apply regression analysis to event log data. Those papers should be analyzed regarding 1) the use case or objective that the technique is used for (e.g., measure, predict, etc.), 2) the element of analysis (e.g., process time, outcome, next activity, failure rate, etc.), and 3) the event log attributes used as input for regression analysis.

Literature:
- Márquez-Chamorro, A. E., Resinas, M., & Ruiz-Cortés, A. (2017). Predictive monitoring of business processes: a survey. IEEE Transactions on Services Computing, 11(6), 962-977.
- Polato, M., Sperduti, A., Burattin, A., & Leoni, M. D. (2018). Time and activity sequence prediction of business process instances. Computing, 100(9), 1005-1031.