Topic 1: Exploratory Process Mining

Exploration is a fundamental phase in many data science disciplines, including process mining. In process mining, exploration typically occurs during the initial analysis of event logs, process discovery, or when unexpected questions arise. Despite its importance, most research has focused on technical aspects, with limited attention to the exploratory phase itself.

The goal of this seminar thesis is to provide a comprehensive understanding of exploratory process mining. The participant should analyze its activities, challenges, and insights based on existing literature. Additionally, the thesis should summarize the current research focus in this area and identify any gaps. If applicable, similarities and differences to exploratory practices in other data science domains can also be discussed.

Relevant Terms:

- Exploration, Human-in-the-loop, Process discovery, Process analysis, Exploratory Process Mining

Topic 2: Probing in Natural Language Processing

Modern NLP models, particularly large pre-trained language models (LMs), achieve impressive performance across a wide range of tasks. However, understanding what these models actually learn about language remains a major research challenge. Probing is a methodology designed to investigate the linguistic knowledge encoded in model representations by training lightweight classifiers (probes) to predict linguistic properties, such as syntax, semantics, or world knowledge.

The goal of this seminar thesis Is to explore probing techniques In NLP and associated fields. Prime interest are the generated insights into model behavior. The participant should investigate:

1. Methodology of Probing: How probes are designed, including linear vs. nonlinear probes, control tasks, and evaluation measures to avoid misleading conclusions.
2. Applications of Probing: Which linguistic phenomena are investigated (e.g., part-of-speech tags, syntactic trees, coreference, semantic roles) and how probing helps understand generalization, model biases, or the evolution of knowledge across layers of a model.
3. Challenges and Limitations: Limitations of probing, such as probe complexity, confounding factors, and debates on whether probing truly reveals model knowledge or merely memorization.

Relevant Terms:

- Probing, Embedding, Natural language processing, Semantic meaning, Representation Learning

Topic 3: Process Mining at the Inter-Organizational Level

Inter-organizational processes have become increasingly relevant, as many real-world business processes — such as supply chains, inter-company workflows, or multi-stakeholder networks — inherently span multiple organizational boundaries and cannot be adequately understood when analyzed within a single enterprise. In this context, inter-organizational process mining has gained importance because it extends traditional process mining by considering event data across organizations. Thus, inter-organizational process mining must integrate event logs from multiple autonomous organizations, addressing challenges such as heterogeneous data sources, varying logging standards, distributed data ownership, and confidentiality or trust constraints.

This seminar thesis aims to investigate and critically assess the current state of research on inter-organizational process mining. Therefore, the main requirements, challenges, and opportunities associated with inter-organizational process mining, as well as the existing methodological approaches and frameworks should be identified. The overall objective is to provide a structured and comprehensive overview of inter-organizational process mining, highlighting its current capabilities, limitations, and potential directions for future research.

Relevant Terms:

- Organization, Process Mining, Inter-organizational process, Event Log, Data Ownership

Topic 4: Process Representation Learning

Process representation learning refers to methods for transforming business process data into structured representations that can be used for machine learning. Instead of working directly with raw event logs or process models, these approaches aim to encode processes in a form that captures relevant behavioral or structural information while remaining computationally manageable. This is relevant because many machine learning applications in process mining depend on how processes are represented. Different representations can emphasize different aspects of a process and influence the results of subsequent analyses. Understanding the basic idea of process representation learning helps explain how process data can be prepared for advanced analysis and why representation choices matter, even before any modeling or prediction takes place.

The goal of this seminar thesis is to provide an overview of the current research on process representation learning. Specifically, the participant should (1) identify the different existing approaches, (2) explicate which aspects of the process they emphasize, and (3) reflect on how this emphasis could influence any subsequent analysis. A special focus should be put on the inclusion of natural language semantics.

Relevant Terms:

- Representation learning, Trace, Encoding, Embedding, Context attributes

Topic 5: Multi-perspective conformance checking

Conformance checking in process mining aims to compare real-life event logs with predefined process models to detect deviations. Traditional approaches often focus on the control-flow perspective—i.e., whether the sequence of activities in the log aligns with the allowed behavior in the model. However, real-world processes are inherently multi-perspective, involving not just control flow but also data, time, and resources. Multi-perspective conformance checking aims to assess alignment between models and logs across these different dimensions simultaneously.

The goal of this seminar thesis is to explore the techniques and frameworks developed for multi-perspective conformance checking. The participant should investigate: (1) Which additional perspectives are typically considered (e.g., data, time, resource); (2) How these perspectives are modeled and aligned with event data (e.g., through data-aware Petri nets, Declare constraints); (3) What types of deviations can be detected (e.g., data inconsistencies, time violations, unauthorized resource usage) and how they are evaluated. The thesis should also discuss trade-offs between completeness, precision, and scalability in multi-perspective checking, and analyze how these methods compare to control-flow-only techniques.

Relevant Terms:

- Process model, Event log, Process perspective, Alignment, Conformance checking