

Prof. Dr. Jana-Rebecca Rehse
L 15, 1-6, Room 413
68161 Mannheim
Germany
Phone +49 621 181-1766
rehse@bwl.uni-mannheim.de

Mining Reference Models from Large Business Process Collections

Reference models play an important role in business process management in that they typically serve as starting points for large business and enterprise system transformation initiatives. However, it can be challenging to find reference models that are suitable for a given transformation scenario. Often, reference models are proprietary knowledge closely kept by specialized consultancies, and hence expensive to obtain and not available for most organizations. Also, assumptions that were made when manually creating a reference model are not sufficiently revised over time, i.e., a given reference model may not reflect the experiences and learnings that were accumulated during the projects that the model served.

To address these shortcomings, it is necessary to inductively define and refine reference models based on model variants that exist in different process landscapes. This can be achieved by searching through large cross-organizational process model collections (hence: process landscape collections), as for example maintained by business process modeling Software-as-a-Service (SaaS) vendors. While there is some academic interest in the mining of reference models from process model collections, little is known about the feasibility of these approaches for practical purposes.

Master Thesis Topic: Identifying Candidates for Reference Model Mining in the Wild

Academic literature on reference model mining from process model collections is relatively sparse, and existing approaches typically focus on control flow properties, i.e., activity orderings. In practice, however, other model properties, such as diagram name, non-graphical diagram-level and element-level attributes, as well as non-control flow elements such as roles/resources (BPMN pools and lanes), documents, and data objects, are just as relevant. In particular, natural language properties are promising candidates for straightforward process variant identification across process landscapes, and roles, documents, and IT system definitions should not come as an afterthought in reference models. This master thesis sets out to explore these practical perspectives, given the academic state-of-the-art.

To this end, we will identify process model properties and algorithms for identifying reference model candidates in a large collection of business processes. Specifically, this includes the following tasks:

- (i) Identify relevant process model properties for clustering models into reference model-oriented groups.
- (ii) Evaluate different clustering algorithms on a real-world dataset¹.

¹ As a starting point, this will most likely be an SAP-internal dataset of the Signavio Academic Initiative, with millions of models, many of which are, however of low quality.

Requirements:

- Hands-on mentality and good communication and collaboration skills
- Experience and strong interest in business process modeling with BPMN
- Experience in data manipulation and data-oriented programming in Python, i.e., pandas, numpy, and scikit-learn and/or
- Substantial real-world software engineering experience
- Ability and willingness to work with messy data
- Basic knowledge in Machine Learning, with a focus on unsupervised methods like clustering
- Willingness to collaborate with process experts to reconcile findings and results
- Good knowledge of software engineering practices and tools, like git and test-driven development approaches

Literature:

- Rehse, J. Hake, P. Fettke, P. Loos, P. "Inductive reference model development: recent results and current challenges" (2016). Informatik 2016. Gesellschaft für Informatik eV.
- Rehse, J. Fettke, P. Loos, P. "A graph-theoretic method for the inductive development of reference process models" (2017). *Software & Systems Modeling*, 16(3), 833-873.
- Rehse, J. Fettke, P. "Clustering business process activities for identifying reference model components" (2018). *International Conference on Business Process Management*. Springer.
- Scholta, H., Niemann, M., Delfmann, P., Räckers, M., & Becker, J. "Semi-automatic inductive construction of reference process models that represent best practices in public administrations: a method" (2019). *Information Systems*, 84, 63-87.