On the detection and prediction of errors in EPC business process models - extended abstract -

Jan Mendling[1,2]

[1]Faculty of Information Technology
Queensland University of Technology
Level 5, 126 Margaret Street, Brisbane QLD 4000, Australia
j.mendling@qut.edu.au
[2] Institute of Information Systems and New Media
Vienna University of Economics and Business Administration
Augasse 2-6, A-1090 Wien, Austria
jan.mendling@wu-wien.ac.at

Abstract: This paper gives an extended abstract of my doctoral thesis on detection and prediction of errors in EPC business process models (see [Men07]).[1] This research is motivated by the observation that business process models are valuable design artifacts in the business process management life cycle, and therefore, they are subject to quality considerations. In this context, the absence of formal errors, such as deadlocks, is of paramount importance for the subsequent implementation of the process. This doctoral thesis provides a fourfold contribution to the understanding of such errors in business process models with a particular focus on Event-driven Process Chains (EPCs), a business process modeling language that is frequently used in practice. Firstly, we formalize the operational semantics of EPCs in a novel way so that matching OR-splits and OR-joins never deadlock. Secondly, and based on these semantics, we introduce a soundness criterion for EPCs that offers a precise identification of those models which have errors. For the verification of this soundness notion in practice, we present two analysis tools, a ProM plug-in for a verification based on the reachability graph, and a batch program called *xoEPC* for a verification based on reduction rules. Thirdly, we define a set of business process model metrics that are supposed to serve as predictors for error probability of an individual EPC. Fourthly, we use statistical methods and a sample of about 2000 EPCs from practice to derive a regression function for the prediction of error probability. This function is validated against a holdout set of 113 EPCs from textbooks showing that 90% of the EPCs could be classified correctly as having errors or not. These results emphasize the importance of quality issues in business process modeling and provides the foundations for innovations in tool support.

[1]A PDF of the thesis can be downloaded at
http://wi.wu-wien.ac.at/home/mendling/publications/Mendling%20Doctoral%20thesis.pdf

1 Introduction

It is a fundamental insight of software engineering that design errors should be detected as early as possible (see e.g. [Boe81, Moo05]). The later errors are detected, the more rework has to be done, and the more design effort has been at least partially useless. This also holds for the consecutive steps of analysis, design, and implementation in the business process management life cycle (cf. e.g. [Ros06]). In the design phase, process models are typically created with semi-formal business process modeling languages while formal executable models are needed for the implementation. This problem is often referred to as the gap between business process design and implementation phase (see e.g. [MR04]). Therefore, quality frameworks, such as the Guidelines of Process Modeling [BRU00], stress correctness as the most important quality attribute of business process models.

In the following sections we briefly sketch the results of this doctoral thesis. In Section 2 we present the research contribution of the thesis. Section 3 discusses the findings and its implications for business process modeling. Finally, Section 4 gives an outlook on future research.

2 Research Contribution

The research objective of this doctoral thesis is the development of a framework for the detection of formal errors in business process models, and the prediction of error probability based on quality attributes of these models. We will focus on Event-driven Process Chains (EPCs), a business process modeling language that is heavily used in practice. The advantage of this focus is, firstly, that the results of this thesis are likely to have an impact on current modeling practices. Secondly, there is a large empirical basis for analysis. By tapping the extensive stock of EPC model collections, we aim to bring forth general insights into the connection between process model attributes and error probability. In order to validate such a connection, we first need to establish an understanding of model attributes that are likely connected with error probability. Furthermore, we must formally define an appropriate notion of correctness, which gives an answer to the question whether a model has a formal error or not. It is a prerequisite to answering this question that we define the operational semantics of the process modeling language, i.e. EPCs, in a formal way. Against the state of the art, this thesis provides the following technical contributions.

Formalization of the OR-join: The semantics of the OR-join have been debated for more than 10 years now. Existing formalizations suffer either from a restriction of the EPC syntax (see e.g. [CS94, LA94, LSW98, Aal99, DR01]) or from non-intuitive behavior (see e.g. [NR02, Kin06, AH05, WEAH05]). In Chapter 2 of this thesis we formalize the EPC semantics concept as proposed in [MA06]. In comparison to other approaches, this novel formalization has the advantage that it is not restricted to a subset of EPCs, and that it provides intuitive semantics for blocks of matching OR-splits and joins since they cannot deadlock. The calculation of the reachability graph was implemented as a plug-in for ProM as a proof of concept. In this way, this novel semantics definition contributes to research on the specification of business process modeling languages.

Verification of process models with OR-joins and multiple start and end events: Verification techniques for process models with OR-joins and multiple start and end events suffer from one of two problems. Firstly, they build on an approximation of the actual behavior and, therefore, do not provide a precise answer to the verification problem, e.g. by considering a relaxed notion of soundness [DR01], by involving user decisions [DAV05], or by approximating relaxed soundness with invariants [VA06]. Secondly, there are verification approaches for semantics definitions (see [CFK05, WAHE06]) that suffer from the previously mentioned non-intuitive behavior. While this is not a problem of the verification itself, all these approaches are not tailored to cope with multiple start and end events. In Chapter 4 of this thesis, we specify a dedicated soundness criterion for EPC business process models with OR-joins and multiple start and end events. Furthermore, we define two verification approaches for EPC soundness, one as an explicit analysis of the reachability graph, and a second based on reduction rules to provide a better verification performance. Both approaches were implemented as a proof of concept. In this way, we contribute to the verification of process models with OR-joins and multiple start and end events, and in particular, we extend the set of reduction rules for business process models.

Metrics for business process models: Metrics play an important role in the operationalization of various quality-related aspects in software engineering, network analysis, and business process modeling. Several authors use metrics to capture different aspects of business process models that are presumably related to quality (see e.g. [LY92, Nis98, Mor99, RV04, Car05, BG05, CGP+05, CMNR06, LG06, ARGP06, MMN+06a, MMN+06b]). A problem of these works is that business process-specific concepts like sequentiality, decision points, concurrency, or repetition are hardly considered, and too often simple count metrics are defined. Furthermore, there appears to be little awareness of related research, maybe because process model measurement is conducted in separate disciplines including software process management, network analysis, Petri nets theory, and conceptual modeling. In Chapter 5 of this thesis, we will provide an extensive list of metrics for business process models and relate it to previously isolated research. Beyond that, we provide a detailed discussion of the rationale and the limitations of each metric, which is meant to serve as a predictor for error probability. We formulate a hypothesis for each metric based on whether it is positively or negatively correlated with error probability.

Validation of metrics as error predictors: Up to now, there is little empirical evidence for the validity of business process model metrics as predictors for error probability. Some empirical work was conducted, but with a different focus. *Lee and Yoon* investigate the empirical relationship between parameters of Petri nets and their state space [LY92]. *Canfora et al.* empirically evaluate the suitability of metrics to serve as predictors for maintainability of the process model [CGP+05]. *Cardoso* analyzes the correlation between the control flow complexity metric with the perceived complexity of process models [Car06]. Most related to this thesis is an analysis of the SAP Reference Model where *Mendling et al.* test a set of simple count metrics as error predictors [MMN$^+$06a, MMN$^+$06b]. In Chapter 6 of this thesis, we use logistic regression for the test which is similar to the analysis of the SAP Reference Model. Still, we consider both the broader set of metrics from Chapter 5, a precise notion of EPC soundness as defined in Chapter 4, and a much broarder sample of EPC models from practice. The results do not only show that certain metrics are indeed a good predictor for error probability, but also that simple count metrics fail to capture important aspects of a process model.

Little research in information systems tries to combine design science and behavioral science research paradigms (see e.g. [BH05]). Since the previously listed contributions cover both design and behavioral aspects, we consider the main contribution of this thesis to be the innovative and holistic combination of both these research paradigms in order to deliver a deeper understanding of errors in business process modeling.

3 Discussion

The results of this thesis have some more general implications for business process modeling. On the following pages we discuss the implications for 1) the importance of verification in practice, 2) for improvements of the business process modeling process, 3) for future business process modeling tools, and 4) for teaching of business process modeling languages.

1. *Importance of Verification:* The amount of errors in the different EPC model collections from practice that we used in this thesis emphasizes the importance of verification. We showed that an error ratio of about 10% is the average over the samples, with 3.3% being the minimum. While verification has been discussed for some time, this thesis demonstrates that the different techniques have matured to handle large sets of several thousand business process models on a common desktop computer. This observation relates to the gap between business analysis and information systems engineering in business process modeling (see [STA05, p.141] or [HR07, pp.424]), i.e. the refusal of engineers to reuse process documentations for systems implementation. While this gap is frequently accepted as a natural breach, this thesis tells a different story: the considerable amount of formal errors in documentation models makes it hardly possible to directly reuse them in the implementation. Therefore, the utilization of verification techniques in practice might be the key to eventually close this gap in the future.

2. *Business Process Modeling Process:* In this thesis, we gathered substantial theoretical and statistical evidence that formal errors are connected with several characteristics of the process model. This finding provides the opportunity to use process model metrics for the management of the design process and of process model quality. This is in particular the case if different design options have to be evaluated, and one

of multiple models might be considered to be superior regarding error-probability based on some metric. Furthermore, the strong connection between size and errors offers objective input for decisions regarding the question when a model should be split up or when model parts should be put in a sub-process. There is clearly a need for further research in this area. Nevertheless, our findings represent a major step towards establishing business process modeling as an engineering discipline beyond the intuition of the modeler.

3. *Business Process Modeling Tools:* Both items 1) and 2) call for respective tool support. While the verification techniques are apparently mature enough to deal with large models from practice, there seems to be too little attention paid to this issue by tool vendors. Indeed, tool vendors should have an interest in these topics since the lack of respective features has a negative impact on the productivity of the business process modeling exercise: models cannot be reused for system development, business users cannot interpret the models properly, and conclusions can hardly be drawn from the models regarding process performance. Beyond verification support, modeling tools could easily calculate process metrics to assist the modeler in the design task. Building on such features, the tool vendors could easily provide a greater benefit to their customers and help to improve the process of designing business process models.

4. *Teaching Process Modeling:* Apparently, there is a weakness in the way business process modeling is taught, if practitioners introduce a formal error in every tenth model (at least in our sample). Even worse, the four textbooks on business process modeling that we used to build the holdout sample had an even higher error ratio. While these rates might be partially attributed to missing verification support in the tools, there seems to be a problem for many modelers to understand the behavioral implications of their design. This has two consequences. Firstly, teaching of business process modeling should pay less attention to teaching a specific business process modeling language, but instead focus on conveying the general principles behind it, i.e., concurrency, synchronization, repetition, and other aspects captured, as by the workflow patterns. Secondly, formal errors seem to get too little attention. Concepts like deadlocks should not only be taught as a technical property of a business process model, but also an erroneous business rule that also leads to problems in real-world business processes that are not supported by information systems. Furthermore, the metrics are a good starting point for teaching patterns that are unlikely to result in errors, such as a high degree of structuredness appears to be less prone to cause errors. Such an approach might eventually deliver a better awareness and attention of formal errors in business process modeling practice.

4 Future Research

There are several open questions that could not be addressed in detail in this thesis. In particular, we focused on business process model metrics and their capability to predict errors in business process models. We found strong evidence that our set of metrics can indeed explain a great share of the variation in error probability. Still, there are other factors we did not investigate in detail including personal factors, modeling purpose, domain knowledge, modeling language, or graphical layout that all might be connected with error probability. Furthermore, they might also be related to other important quality aspects that we did not analyze, like maintainability or understandability. In particular, we strongly agree with *Moody* [Moo05] who calls for more empirical research in business process modeling. This thesis and its findings gives an idea of the benefits we might gain from this research, and therefore may be regarded as an encouragement to follow *Moody's* call.

References

[Aal99] W.M.P. van der Aalst. Formalization and Verification of Event-driven Process Chains. *Information and Software Technology*, 41(10):639–650, 1999.

[AH05] W.M.P. van der Aalst and A.H.M. ter Hofstede. YAWL: Yet Another Workflow Language. *Information Systems*, 30(4):245–275, 2005.

[ARGP06] E. Rolón Aguilar, F. Ruiz, F. García and M. Piattini. Towards a Suite of Metrics for Business Process Models in BPMN. In Y. Manolopoulos, J. Filipe, P. Constantopoulos and J. Cordeiro, editors, *ICEIS 2006 - Proceedings of the Eighth International Conference on Enterprise Information Systems: Databases and Information Systems Integration (III), Paphos, Cyprus, May 23-27, 2006*, pages 440–443, 2006.

[BG05] S. Balasubramanian and M. Gupta. Structural metrics for goal based business process design and evaluation. *Business Process Management Journal*, 11(6):680–694, 2005.

[BH05] H.U. Buhl and B. Heinrich. Meinung/Dialog: Empirical Research Strategies in Conceptual Modeling - Silver Bullet or Academic Toys? *Wirtschaftsinformatik*, 47(2):152–162, 2005.

[Boe81] B.W. Boehm. *Software Engineering Economics*. Prentice-Hall, Englewood Cliffs, 1981.

[BRU00] J. Becker, M. Rosemann and C. von Uthmann. Guidelines of Business Process Modeling. In W.M.P. van der Aalst, J. Desel and A. Oberweis, editors, *Business Process Management. Models, Techniques, and Empirical Studies*, pages 30–49. Springer, Berlin et al., 2000.

[Car05] J. Cardoso. *Workflow Handbook 2005*, chapter Evaluating Workflows and Web Process Complexity, pages 284–290. Future Strategies, Inc., Lighthouse Point, FL, USA, 2005.

[Car06] J. Cardoso. Process control-flow complexity metric: An empirical validation. In *Proceedings of IEEE International Conference on Services Computing (IEEE SCC 06), Chicago, USA, September 18-22*, pages 167–173. IEEE Computer Society, 2006.

[CFK05] N. Cuntz, J. Freiheit and E. Kindler. On the semantics of EPCs: Faster calculation for EPCs with small state spaces. In M. Nüttgens and F.J. Rump, editor, *Proceedings of the 4th GI Workshop on Business Process Management with Event-Driven Process Chains (EPK 2005)*, pages 7–23, Hamburg, Germany, December 2005. German Informatics Society.

[CGP+05] G. Canfora, F. García, M. Piattini, F. Ruiz and C.A. Visaggio. A family of experiments to validate metrics for software process models. *Journal of Systems and Software*, 77(2):113–129, 2005.

[CMNR06] J. Cardoso, J. Mendling, G. Neumann and H. Reijers. A Discourse on Complexity of Process Models. In Johann Eder and Schahram Dustdar, editors, *Proceedings of BPM Workshops 2006*, volume 4103 of *Lecture Notes in Computer Science*, pages 115–126, Vienna, Austria, 2006. Springer-Verlag.

[CS94] R. Chen and A. W. Scheer. Modellierung von Prozessketten mittels Petri-Netz-Theorie. Heft 107, Institut für Wirtschaftsinformatik, Saarbrücken, Germany, 1994.

[DAV05] B.F. van Dongen, W.M.P. van der Aalst and H.M.W. Verbeek. Verification of EPCs: Using Reduction Rules and Petri Nets. In O. Pastor and J. Falcão e Cunha, editors, *Advanced Information Systems Engineering, 17th International Conference, CAiSE 2005, Porto, Portugal, June 13-17, 2005, Proceedings*, volume 3520 of *Lecture Notes in Computer Science*, pages 372–386. Springer, 2005.

[DR01] J. Dehnert and P. Rittgen. Relaxed Soundness of Business Processes. In K.R. Dittrick, A. Geppert and M.C. Norrie, editors, *Proceedings of the 13th International Conference on Advanced Information Systems Engineering*, volume 2068 of *Lecture Notes in Computer Science*, pages 151–170, Interlaken, 2001. Springer.

[HR07] M. Hepp and D. Roman. An Ontology Framework for Semantic Business Process Management. In A. Oberweis, C. Weinhardt, H. Gimpel, A. Koschmider, V. Pankratius and B. Schnizler, editors, *eOrganisation: Service-, Prozess-, Market Engineering*, pages 423–440. Universitätsverlag Karlsruhe, 2007. Tagungsband der 8. Internationalen Tagung Wirtschaftsinformatik. Band 1.

[Kin06] E. Kindler. On the semantics of EPCs: Resolving the vicious circle. *Data & Knowledge Engineering*, 56(1):23–40, 2006.

[LA94] F. Leymann and W. Altenhuber. Managing Business Processes as an Information Resource. *IBM Systems Journal*, 33(2):326–348, 1994.

[LG06] R. Laue and V. Gruhn. Complexity Metrics for Business Process Models. In Witold Abramowicz and Heinrich C. Mayr, editors, *9th International Conference on Business Information Systems (BIS 2006)*, volume 85 of *Lecture Notes in Informatics*, pages 1–12, 2006.

[LSW98] P. Langner, C. Schneider and J. Wehler. Petri Net Based Certification of Event driven Process Chains. In J. Desel and M. Silva, editor, *Application and Theory of Petri Nets*, volume 1420 of *Lecture Notes in Computer Science*, pages 286–305, 1998.

[LY92] G.S. Lee and J.-M. Yoon. An empirical study on the complexity metrics of Petri nets. *Microelectronics and Reliability*, 32(3):323–329, 1992.

[MA06] J. Mendling and W.M.P. van der Aalst. Towards EPC Semantics based on State and Context. In M. Nüttgens and F.J. Rump and J. Mendling, editor, *Proceedings of the 5th GI Workshop on Business Process Management with Event-Driven Process Chains (EPK 2006)*, pages 25–48, Vienna, Austria, December 2006. German Informatics Society.

[Men07] J. Mendling. *Detection and Prediction of Errors in EPC Business Process Models*. PhD thesis, Vienna University of Economics and Business Administration, 2007.

[MMN+06a] J. Mendling, M. Moser, G. Neumann, H.M.W. Verbeek, B.F. van Dongen and W.M.P. van der Aalst. A Quantitative Analysis of Faulty EPCs in the SAP Reference Model. BPM Center Report BPM-06-08, BPMCenter.org, 2006.

[MMN+06b] J. Mendling, M. Moser, G. Neumann, H.M.W. Verbeek, B.F. van Dongen and W.M.P. van der Aalst. Faulty EPCs in the SAP Reference Model. In J.L. Fiadeiro S. Dustdar and A. Sheth, editors, *Proceedings of BPM 2006*, volume 4102 of *Lecture Notes in Computer Science*, page 451457, Vienna, Austria, 2006. Springer-Verlag.

[Moo05] D.L. Moody. Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions. *Data & Knowledge Engineering*, 55(3):243–276, 2005.

[Mor99] S. Morasca. Measuring Attributes of Concurrent Software Specifications in Petri Nets. In *METRICS '99: Proceedings of the 6th International Symposium on Software Metrics*, pages 100–110, Washington, DC, USA, 1999. IEEE Computer Society.

[MR04] M. zur Muehlen and M. Rosemann. Multi-Paradigm Process Management. In *Proc. of the Fifth Workshop on Business Process Modeling, Development, and Support - CAiSE Workshops*, 2004.

[Nis98] M.E. Nissen. Redesigning reengineering through measurement-driven inference. *MIS Quarterly*, 22(4):509–534, 1998.

[NR02] M. Nüttgens and F.J. Rump. Syntax und Semantik Ereignisgesteuerter Prozessketten (EPK). In J. Desel and M. Weske, editor, *Proceedings of Promise 2002, Potsdam, Germany*, volume 21 of *Lecture Notes in Informatics*, pages 64–77, 2002.

[Ros06] M. Rosemann. Potential pitfalls of process modeling: part A. *Business Process Management Journal*, 12(2):249–254, 2006.

[RV04] H.A. Reijers and I.T.P. Vanderfeesten. Cohesion and Coupling Metrics for Workflow Process Design. In J. Desel, B. Pernici and M. Weske, editors, *Business Process Management: Second International Conference, BPM 2004, Potsdam, Germany, June 17-18, 2004. Proceedings*, volume 3080 of *Lecture Notes in Computer Science*, pages 290–305. Springer, 2004.

[STA05] A.-W. Scheer, O. Thomas and O. Adam. *Process Aware Information Systems: Bridging People and Software Through Process Technology*, chapter Process Modeling Using Event-Driven Process Chains, pages 119–146. Wiley Publishing, 2005.

[VA06] H. M. W. Verbeek and W.M.P. van der Aalst. On the verification of EPCs using T-invariants. BPMCenter Report BPM-06-05, BPMcenter.org, 2006.

[WAHE06] M.T. Wynn, W.M.P. van der Aalst, A.H.M. ter Hofstede and D. Edmond. Verifying Workflows with Cancellation Regions and OR-joins: An Approach Based on Reset Nets and Reachability Analysis. In S. Dustdar, J.L. Fiadeiro, and A. Sheth, editors, *Proceedings of BPM 2006*, volume 4102 of *Lecture Notes in Computer Science*, pages 389–394, Vienna, Austria, 2006. Springer-Verlag.

[WEAH05] M.T. Wynn, D. Edmond, W.M.P. van der Aalst and A.H.M. ter Hofstede. Achieving a General, Formal and Decidable Approach to the OR-join in Workflow using Reset nets. In G. Ciardo and P. Darondeau, editors, *Applications and Theory of Petri Nets 2005*, volume 3536 of *Lecture Notes in Computer Science*, pages 423–443, 2005.