Preface

Business process modeling plays an important role in the management of business processes. As valuable design artifacts, business process models are subject to quality considerations. The absence of formal errors such as deadlocks is of paramount importance for the subsequent implementation of the process. This book develops a framework for the detection of formal errors in business process models and for the prediction of error probability based on quality attributes of these models (metrics). We focus on Event-driven Process Chains (EPCs), a widely used business process modeling language due to its extensive tool support. The advantage of this focus is firstly that the results of this book can be directly translated into process modeling practice. Secondly, there is a large empirical basis of models. By utilizing this large stock of EPC model collections, we aim to bring forth general insights into the connection between process model metrics and error probability. In order to validate such a connection, we first need to establish an understanding of which model attributes are likely connected with error probability. Furthermore, we must formally define an appropriate notion of correctness that answers the question of whether or not a model has a formal error. As a prerequisite to answering this question, we must define the operational semantics of the process modeling language formally.

Contributions

This book presents a precise description of EPCs, their control-flow semantics and a suitable correctness criterion called EPC soundness. Furthermore, we identify theoretical arguments on why structural metrics should be connected with error probability and provide an empirical validation of this connection. To be more concise, this book provides the following technical contributions.

Formalization of the OR-join: The semantics of the OR-join have been debated for more than a decade. Existing formalizations suffer from either a restriction of the EPC syntax (see [78, 247, 238, 4, 101]) or from non-intuitive behavior (see [325, 218, 11, 465]). In Chap. 2, we formalize the EPC semantics concept as proposed elsewhere [267]. In comparison to other approaches this novel formalization has the advantage of not being restricted to a subset of EPCs. Moreover, it provides intuitive semantics for blocks of matching OR-splits and joins since they cannot deadlock. As a proof of concept, we implemented a plug-in for ProM that calculates the reachability graph. 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, e.g., by considering a relaxed notion of soundness [101], by involving user decisions [109] or by approximating relaxed soundness with invariants [440]. Therefore, they do not provide a precise answer to the verification problem. Secondly, some verification approaches for semantics definitions (see [88, 464]) suffer from the previously mentioned non-intuitive behavior. While this is not the result of the verification problem itself, none of these approaches has been tailored to cope with multiple start and end events. In Chap. 3, we specify a dedicated soundness criterion for EPC business process models with OR-joins and multiple start and end events. We also 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. Importantly, we also 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 [244, 320, 308, 348, 72, 37, 67, 74, 241, 356, 275, 276]). Unfortunately, business process-specific concepts such as sequentiality, decision points, concurrency, and repetition are hardly considered while simple count metrics are often defined. There also appears to be little awareness of related research, possibly owing to the fact that process model measurement is conducted in separate disciplines such as software process management, network analysis, Petri nets theory, and conceptual modeling. In Chap. 4, we provide an extensive list of metrics for business process models and provide links to previously isolated research. Beyond that, we provide a detailed discussion of the rationale and limitations of each metric 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: Until now, there has been little empirical evidence for the validity of business process model metrics as predictors for error probability. Some empirical work has been conducted; however, it has always maintained a different focus: *Lee and Yoon* investigate the empirical relationship between parameters of Petri nets and their state space [243, 244]. *Canfora et al.* empirically evaluate the suitability of metrics to serve as predictors for maintainability of the process model [67]. *Cardoso* analyzes the correlation between the control flow complexity metric with the perceived complexity of process models [73]. Of most significance to this book is an analysis of the SAP Reference Model in which *Mendling et al.* test a set of simple count metrics as error predictors [275, 276]. In Chap. 5 we use logistic regression for the test, which is similar to the analysis of the SAP Reference Model. We consider both the broader set of metrics from Chap. 4, a precise notion of EPC soundness as defined in Chap. 3, and a much broader sample of EPC models from practice. The results show not only 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.

So little research on information systems and conceptual modeling combines design science and behavioral science research paradigms that there is clearly a need for more empirical insight [306]. Since the previously listed contributions cover both design and behavioral aspects, we consider the main contribution of this book to be the innovative and holistic combination of both these research paradigms in an effort to deliver a deeper understanding of errors in business process modeling.

Structure

This book is organized in six chapters. Beginning with a general overview of business process management, we continue with semantics of EPCs and the verification of soundness before discussing metrics for business process models which are subsequently validated for their capability to predict error probability.

Chapter 1 – *Business Process Management:* In this chapter, we discuss the backgrounds of business process management and define important related terms. We also sketch the importance of business process modeling and the role of errors in the business process management lifecycle.

Chapter 2 – *Event-driven Process Chains (EPC):* This chapter gathers state-of-the-art work on EPCs. Building on the foundations of prior work, we establish a novel syntax definition and a novel semantics definition for EPCs. Our semantics are based on transition relations that define both state changes and context changes. We then present an algorithm to calculate the reachability graph of an EPC based on the transition relations and a respective implementation as a plug-in for ProM. The major motivations for these novel semantics are semantic gaps and non-intuitive behavior of existing formalizations.

Chapter 3 – *Verification of EPC Soundness:* This chapter presents an EPC-specific version of soundness as a criterion of correctness for EPCs. We propose two different approaches for the verification of soundness: one based on the reachability graph and another based on reduction rules. While the first approach explicitly considers all states and transitions that are represented by an EPC, there is a problem with state explosion due to the maximum number of states growing exponentially with the number of arcs. In order to avoid a performance problem we introduce a set of reduction rules. This set extends prior work with new reductions for start and end components, delta components, prism components and homogeneous EPCs. This approach is tested by reducing the SAP Reference Model and shows that the reduction is *fast*, provides a *precise* result for almost all models, and finds *three times as many errors* as other approaches based on relaxed soundness.

Chapter 4 – *Metrics for Business Process Models:* This chapter discusses the suitability of business process model metrics predicting error probability from a theoretical point of view. Revisiting related research in the area of network analysis, software measurement, and metrics for business process models, we find that several aspects of process models have not yet been combined in an overall measurement framework. Based on theoretical considerations we present a set of 15 metrics related to size and 13 metrics that capture various aspects of the structure and the state space of the process model. For each of the metrics we discuss their presumable connection with error probability and formulate respective hypotheses.

Chapter 5 – *Validation of Metrics as Error Predictors:* In this chapter, we conduct several statistical analyzes related to the connection of metrics and error probability. The results of the correlation analysis and the logistic regression model strongly confirm the hypothetical impact direction of the metrics. We then derive a logistic regression function, based on a sample of approximately 2000 real EPC business process models, that correctly classifies 90% of the models from a second independent sample.

Chapter 6 – *Implications for Business Process Modeling:* Here we present a summary of the findings and offer an outlook on future research. A major result is a set of seven guidelines of process modeling. Beyond that, we discuss the implications for the business process modeling process, respective tool support, EPCs as a business process modeling language, and teaching of business process modeling.

Acknowledgements

The major share of this book stems directly from my doctoral thesis, submitted to the Vienna University of Economics and Business Administration (WU Wien) in May 2007. I would like to thank my supervisors, Prof. Dr. Gustaf Neumann and Prof. Dr. Wil van der Aalst. In October 2003 I began working at the Institute of Information Systems and New Media. Prof. Neumann provided me with a creative work environment where I could further develop my ideas and academic writing skills. He was open and supportive in all matters, particularly when it came to modifying the topic of my thesis. In 2004 I met Prof. van der Aalst at a conference and he agreed to be my second supervisor. His feedback was extraordinarily helpful for the formalization of my concepts.

Additionally, I would like to thank Prof. Dr. Markus Nüttgens for fueling my interest in business process modeling and EPCs, and for his continuous support since my diploma thesis in Trier, Germany; Prof. Dr. Walter Schertler and Prof. Dr. Axel Schwickert for serving as references for my application to WU in Vienna; Prof. Dr.

Carlos de Backer for being a role model in teaching information systems and explaining IT concepts in Dutch; and Prof. Dr. Hans Czap for his efforts in establishing the Diplomstudiengang Wirtschaftsinformatik at the University of Trier. I would like to thank all my former colleagues at the Institute of Information Systems and New Media and the Institute of Management Information Systems at the Vienna University of Economics and Business Administration (WU Wien) for providing a friendly work environment. Finally, thanks to Prof. Dr. Michael Rosemann and A/Prof. Dr. Arthur ter Hofstede and all my colleagues at the BPM group at the Queensland University of Technology Brisbane for their inspirations.

I was happy to discuss research and to collaborate on different paper projects with several excellent people: Wil van der Aalst, Paul Barborka, Alberto Brabenetz, Jan vom Brocke, Jorge Cardoso, Malu Castellanos, Remco Dijkman, Boudewijn van Dongen, Marlon Dumas, Rik Eshuis, Florian Gottschalk, Michael Genrich, Michael Hafner, Lukas Helm, Mitra Heravizadeh, Jörg Hoffmann, Arthur ter Hofstede, Thomas Hornung, Cirano Iochpe, Alex Kokkonen, Georg Köldorfer, Agnes Koschmider, Marcello La Rosa, Kristian Bisgaard Lassen, Jean Michael Lau, Ralf Laue, Ana Karla Alves de Medeiros, Joachim Melcher, Jürgen Moormann, Michael Moser, Michael zur Muehlen, Martin Müller, Gustaf Neumann, Jörg Nitzsche, Markus Nüttgens, Cristian Pérez de Laborda, Karsten Plößer, Andreas Pinterits, Adrian Price, Michael Rausch, Jan Recker, Manfred Reichert, Hajo Reijers, Michael Rosemann, Frank Rump, Bernd Simon, Carlo Simon, Guido Sommer, Mark Strembeck, Gerald Stermsek, Lucinéia Heloisa Thom, Roger Tregear, Irene Vanderfeesten, Eric Verbeek, Barbara Weber, Ingo Weber, Ton Weijters, Fridolin Wild, Uwe Zdun, and Jörg Ziemann. Thank you for sharing your insights with me. I also thank Kalina Chivarova for gathering some of the EPC models of the holdout sample in her bachelor thesis. A special thanks goes to Herbert Liebl, who had the nice idea of generating SVG graphics from the error analysis and highlighting the errors in the EPC models. I would also like to thank Jesse Holt for proof-reading the manuscript. Finally, thanks to Yoseba Peña, who coined the title of my thesis.

July 2008 Jan Mendling

http://www.springer.com/978-3-540-89223-6

Metrics for Process Models
Empirical Foundations of Verification, Error Prediction,
and Guidelines for Correctness
Mendling, J.
2008, XX, 194 p., Softcover
ISBN: 978-3-540-89223-6