Preface

The interest in *data science* is rapidly growing. Many consider data science as *the* profession of the future. Just like computer science emerged as a discipline in the 1970s, we now witness the rapid creation of research centers and bachelor/master programs in data science. The hype related to Big Data and predictive analytics illustrates this. Data ("Big" or "small") are essential for people and organizations and their importance will only increase. However, it is not sufficient to focus on data storage and data analysis. A data scientist also needs to relate data to operational processes and be able to ask the right questions. This requires an understanding of end-to-end processes. *Process mining* bridges the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining. Process mining provides a new means to improve processes in a variety of application domains. The omnipresence of event data combined with process mining allows organizations to diagnose problems based on facts rather than fiction.

Although traditional Business Process Management (BPM) and Business Intelligence (BI) technologies received lots of attention, they did not live up to the expectations raised by academics, consultants, and software vendors. Probably, the same will happen to most of the Big Data technologies vigorously promoted today. The goal should be to improve the operational processes themselves rather than the artifacts (models, data, and systems) they use. As will be demonstrated in this book, there are novel ways to put "data science in action" and improve processes based on the data they generate.

Process mining is an emerging discipline providing comprehensive sets of tools to provide fact-based insights and to support process improvements. This new discipline builds on process model-driven approaches and data mining. However, process mining is much more than an amalgamation of existing approaches. For example, existing data mining techniques are too data-centric to provide a comprehensive understanding of the end-to-end processes in an organization. BI tools focus on simple dashboards and reporting rather than clear-cut business process insights. BPM suites heavily rely on experts modeling idealized to-be processes and do not help the stakeholders to understand the as-is processes.

This book presents a range of process mining techniques that help organizations to uncover their actual business processes. Process mining is not limited to process discovery. By tightly coupling event data and process models, it is possible to check conformance, detect deviations, predict delays, support decision making, and recommend process redesigns. Process mining breathes life into otherwise static process models and puts today's massive data volumes in a process context. Hence, managements trends related to process improvement (e.g., Six Sigma, TQM, CPI, and CPM) and compliance (SOX, Basel II, etc.) can benefit from process mining.

Process mining, as described in this book, emerged in the last decade [156, 160]. However, the roots date back about half a century. For example, Anil Nerode presented an approach to synthesize finite-state machines from example traces in 1958 [108], Carl Adam Petri introduced the first modeling language adequately capturing concurrency in 1962 [111], and Mark Gold was the first to systematically explore different notions of learnability in 1967 [61]. When data mining started to flourish in the 1990s, little attention was given to processes. Moreover, only recently event logs have become omnipresent thus enabling end-to-end process discovery. Since the first survey on process mining in 2003 [156], progress has been spectacular. Process mining techniques have become mature and supported by various tools. Moreover, whereas initially the primary focus was on process discovery, the process mining spectrum has broadened markedly. For instance, conformance checking, multi-perspective process mining, and operational support have become integral parts of ProM, one of the leading process mining tools.

The book provides a comprehensive overview of the state-of-the-art in process mining. It is intended as an introduction to the topic for practitioners, students, and academics. On the one hand, the book is accessible for people that are new to the topic. On the other hand, the book does not avoid explaining important concepts on a rigorous manner. The book aims to be self-contained while covering the entire process mining spectrum from process discovery to operational support. Therefore, it also serves as a reference handbook for people dealing with BPM or BI on a day-to-day basis.

The first edition of this book appeared in 2011 under the title "Process Mining: Discovery, Conformance and Enhancement of Business Processes" [140]. Given the rapid developments in process mining, there was a clear need for an updated version. The original book has been extended in several ways. First of all, process mining has been put into the broader context of *data science* (see the new Chap. 1). This explains the new subtitle "Data Science in Action". There is an urgent need for data scientists able to help organizations improve their operational processes. Therefore, the new edition of the book positions process mining in this broader context and relates it to statistics, data mining, Big Data, etc. Second, there has been significant progress in process discovery in recent years. This is exemplified by the family of *inductive mining* techniques that can handle large incomplete event logs with infrequent behavior, but still provide formal guarantees. The basic elements of inductive mining (Sect. 7.5) and the notion of process trees (Sect. 3.2.8) have been added to this book. Third, the notion of *alignments* has become a key concept to relate observed behavior and modeled behavior. The chapter on conformance checking has been extended to carefully introduce alignments (Sect. 8.3). Moreover, next

to fitness, also quality dimensions like *precision* are now defined. Fourth, a chapter on "process mining in the large" (Chap. 12) has been added to illustrate that process mining can exploit modern infrastructures and that process discovery and conformance checking can be decomposed and distributed. Since the first edition of the book, many new process mining products emerged (often inspired by the open source platform ProM and the previous edition of this book). The chapter on tools (Chap. 11) has been completely rewritten and discusses commercial tools like Celonis Process Mining, Disco, Enterprise Discovery Suite, Interstage Business Process Manager Analytics, Minit, myInvenio, Perceptive Process Mining, QPR ProcessAnalyzer, Rialto Process, SNP Business Process Analysis, and webMethods Process Performance Manager (next to open-source initiatives like ProM and RapidProM). Finally, pointers to recent literature have been added and a new section of data quality has been added (Sect. 5.4). These changes justify a revised edition of the book.

The reader can immediately put process mining into practice due to the applicability of the techniques, the availability of (open-source) process mining software, and the abundance of event data in today's information systems. I sincerely hope that you enjoy reading this book and start using some of the amazing process mining techniques available today.

Eindhoven, The Netherlands Wil van der Aalst
January 2016

http://www.springer.com/978-3-662-49850-7

Process Mining
Data Science in Action
van der Aalst, W.M.P.
2016, XIX, 467 p. 250 illus., 13 illus. in color., Hardcover
ISBN: 978-3-662-49850-7