Process Miner – A Tool for Mining Process Schemes from Event-Based Data

Guido Schimm

Oldenburger Forschungs- und Entwicklungsinstitut
für Informatik-Werkzeuge und -Systeme (OFFIS)
schimm@offis.de

Abstract. Today, process schemes are required for a lot of purposes. Extracting process schemes from event-based data is an alternative to creating them manually. Process Miner is a research prototype that can extract process schemes from event-based data. Its extracting procedure is a multistage data mining that uses a special process model. This paper outlines the main features of the tool and gives an insight into the theoretical background. Also, it describes shortly its implementation and outlines its experimental evaluation.

1 Introduction

Execution of processes is often driven by schemes. Most business processes, for example, are executed by many people together with various computer systems and possibly other tools. Here, a process schema acts as a template for process execution. It explicates a specific set of tasks, their order of execution, decision points branching to alternative executing paths, and the rules that are applied to make such decisions. In order to recognize, manage or control processes the respective schemes are needed. For example, a company’s knowledge management saves and propagates process based knowledge in form of process schemes, advisory systems are based on process schemes, and workflow systems need process schemes in order to control business processes automatically.

The implicated need of process schemes requires schema development in any form. In the case that schema development is performed manually, it was realized that it is often difficult, time consuming, error prone, and expensive. An alternative to this is to generate schemes automatically. One kind of generating process schemes is extracting process schemes from event traces of properly executed processes.

Process miner is a tool that implements a complete mining procedure that extracts process schemes from event-based data. The mined schemes are complete and minimal: Complete in the sense that all recorded processes are covered by the extracted schema, minimal in the sense that only recorded processes are covered. Additionally, decision rules needed to decide between alternative paths of execution are extracted by the tool’s mining procedure, too.
2 Process Miner’s Major Features

The main feature of Process Miner is the process mining procedure. Mining process schemes is based on a large amount of appropriate data. Therefore, the tool expects data to be stored in a database. The user starts the procedure from a menu entry of the tool’s graphical user interface. At first, Process Miner connects to a database and shows a dialog that contains a list of all process data that are in the database. The user selects the process data that should be used as input for the mining procedure. The procedure can be controlled through parameter settings. The parameters determine which mining steps should be performed, some adjustable thresholds and the levels of detail for writing the logs. After the user confirms the settings the mining procedure starts. Process Miner automatically performs the procedure as a sequence of several steps. While the mining procedure runs, the user gets a detailed log of all mining steps. The last step represents the extracted schema in a separate editor window.

Beside the mining procedure implementation Process Miner has some other noteworthy features. One of them is the schema editor. It provides two different views of a process schema at the same time. One view represents a schema as diagram, the other view shows a schema in form of a tree. Both views can be scrolled and zoomed independently. Browsing through a schema and analyzing it is very comfortable, in the case that, for example, the tree view represents an overview and the diagram view is used to show details of the schema.

The user can also edit schemes. For this purpose the editor provides functions like add, delete, copy, paste, clone, move etc. that can be applied to schema elements in both views. For example, the user can cut off parts of a mined schema and then build sub schemes or integrate them in other schemes. Also the user can model complete schemes from scratch. All schemes can be exported to other tools in form of text files in XML-format.

Another useful feature of Process Miner is its simulation component. It allows simulating processes from a particular schema in a specific context. Such a simulation context consists of a set of feature-value pairs. According to the current values alternative paths of execution inside a schema are selected by the simulation component. Altering the context of a simulation leads to different process executions. Contexts can be created by the user with an editor. Alternatively, contexts can be created automatically by Process Miner, too. During a simulation all events of starting or stopping tasks and the context states are stored into a database.

3 Implementation

Process Miner is a research prototype that is considered to be a starting point of application-specific versions. The software is designed in an object-oriented manner and implemented with Java 1.3. The current version comes in form of a single jar archive. It consists of approximately 200 classes spread over nine packages.