

Class 21

Learning Objectives

- Understand how process mining is done
- Understand the types of information that process mining can discover
- Be able to mine process information using YAWL and ProM and answer questions about process execution

Readings

This chapter deals with process mining and process simulation. Both are important when trying to improve the efficiency of existing business processes or to try to identify how new processes might behave, given either past information or assumed simulation conditions. The readings for this class (Chapter 17 of the textbook) are not long, and focus on the use of the ProM process mining software in conjunction with YAWL. The simulation software discussed in the textbook, CPNTools, is not available for all operating systems, and we will therefore not use it for this course. However, you should still read those sections to get an idea of how process mining and process simulation can work together.

Chapter 17 (Sections 17.1 to 17.3)

At this point, you will likely have produced different process models in YAWL and executed some of them. Typical questions that often arise in practice now concern the actual process behaviour. We would like to see how it compares to our plans or expectations, how we can improve the process, and discover other interesting bits about the process. Because a workflow engine records all events for all process instances, there is a rich source of information about the actual behaviour of process instances and work items. Some simple questions might be: Which tasks are executed most often? How long do certain tasks take on average? How busy are the people in the organization on average? Who works with whom on process instances. These questions concern three separate issues. First, the control flow is concerned with performance, timing of tasks, probabilities of task execution, etc. This information is mainly used to improve the speed of the overall process. Second, the resource perspective tells us whether the resources that we allocate and use in the process are too few, just right, or too many. If there are too few, work items have to wait unduly; if there are too many, some resources are idle and could be used otherwise.

A comprehensive look at all the different analysis methods for processes that are provided by ProM is beyond the scope of this one class (and indeed beyond the scope of the textbook chapter). What you should take away from this class are some of the simpler kinds of analysis that can be performed on process logs and how to use them to identify problems in a workflow. The textbook briefly describes three examples in Section 17.3; the basic log statistics, the dotted-chart analysis, the LTL checker. You can find these (and many more) in the analysis menu of the ProM software (see below).

As you read the material you might want to reflect on the following: The book (pg. 440) suggests that “in practice, there is often a significant gap between what is prescribed or supposed to happen and what actually happens”. Given that workflow management systems are designed to ensure a consistent outcome that conforms to all business rules, why do you think there is discrepancy. Can you think of ways in which this discrepancy can arise, despite the best efforts of the workflow management system?

Review Questions

- Explain why process mining might be necessary even when the workflow model can be statically analyzed.
- Explain why process mining might be necessary even when the workflow model can be simulated.
- Explain the relationship between process mining and process simulation.
- What three types of information are required for realistic simulation and how are they obtained?
- Explain the purpose and outcomes of performance analysis in process mining
- Explain the purpose and outcomes of resource analysis in process mining
- Explain the purpose and outcomes of conformance analysis in process mining