Kiel University
Christian-Albrechts-Universität zu Kiel
Faculty of Engineering

DEPARTMENT OF COMPUTER SCIENCE

Process Analytics

Prof. Dr. Agnes Koschmider
Dominik Janssen

Advanced Process Mining

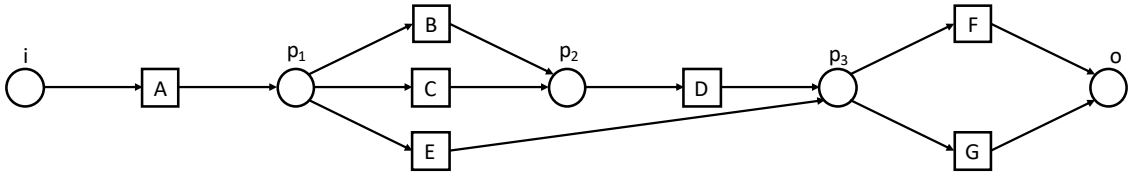
Summer term 2020

Exercise sheet 6

Revision

Exercise 1: Four dimensions of quality

a) Given the following process model,



generate an event log with the least possible number of entries that will result in:

i) Precision = 1

ii) Fitness = 1

b) Calculate the F1-Score for the given values:

Fitness = 0.8

Precision = 0.95

c) Name two measurements to reduce the complexity of real-life logs.

Solution

a) i) For a perfect precision the event log has to contain every possible trace reproducible by the model:

#	Trace
n	ABDF
m	ABDG
o	ACDF
p	ACDG
q	AEF
r	AEG

ii) For a perfect fitness the event log has to contain only one single entry, that can be reproduced exactly by the model:

#	Trace
n	AEG

b) $F1\ Score = \frac{2 \times fitness \times precision}{fitness + precision} = \frac{2 \times 0.8 \times 0.95}{0.9 + 0.95} = 0.8686$

c)
- Abstraction: Only most frequent arcs/activities
- Filtering: Remove events that fulfil a certain condition

Exercise 2: Directly-Follows Graph

a) Name one possible problem with Directly-Follows graphs.

b) True or False: The precision of a Directly-Follows graph is very high.

Solution

a) Possible problems with DFGs are:

- Activities that have a flexible ordering (e.g., due to concurrency) lead to Spaghetti-like DFGs with loops even when activities are executed at most once.
- DFGs can be simplified using frequency-based thresholds. However, this may lead to all kinds of interpretation problems due to ïnvisible gapsïn the model.
- Performance information mapped onto DFGs can be misleading, e.g., the average time reported between two activities is conditional (only the situations where they directly follow each other are considered)

b) **False.** A DFG allows for more behaviour than was recorded in the log.

Exercise 3: Conformance Checking

a) True or False: The effort to find the optimal alignment is independent from the trace length and model size.

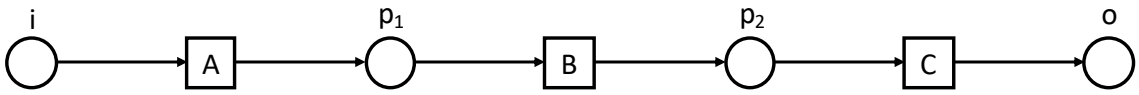
b) True or False: There are always multiple optimal alignments for a given trace and model if the trace is not exactly representable by the model.

c) Name two reasons why conformance checking can be useful.

Solution

a) **False.** The size of the search space is determined by the length of the trace and the size of the model. The longer a trace and the more complicated a process model, the more combinations of alignments have to be considered.

b) **False.** Counterexample:



σ	$\gg$	B	$\gg$
N	A	B	C

c)
- Auditing can be a reason. It helps the auditor to detect fraud or malpractice.
- Improving of the process. Risks and inefficiencies can be detected with conformance checking.