First name	Name	Matr.-Nr.

RWTH Aachen
Lehrstuhl für Informatik 9
Prof. Dr. van der Aalst

Exam 1 - 25.08.2021
Business Process Intelligence

Study course:
□ Diplom Informatik □ Master Informatik □ Other: _____
□ Master SSE □ Master Media Informatics

- Duration of the exam: **120 minutes**
- Write your first name, last name, and student ID (Matrikelnummer) on each sheet.
- Sign this cover sheet.
- Give your solutions in a readable and traceable manner. **Solutions will be graded based on completeness and correctness of the description/application of the algorithm/method**.
- Give your solutions on the exam sheets only. If you need extra paper, use only the paper provided by the supervisors. Make sure to specify your name and Matrikelnummer on all papers.
- Please cross out those things you do not wish to be graded. In the case of multiple answers, the first one will be graded.
- In case of attempted deception, your exam will be graded as **failed**.
- At the end of the exam, hand in your complete copy. Do not separate any sheets by removing the staples.
- This exam accounts for 60% of the grade. The other 40% could have been obtained through the mandatory assignment (you need at least 50% for both the exam and the assignment to pass the course).
- **You may only use a black or blue pen (NOT a pencil), and a basic scientific calculator (no memory or automatic plotting); no additional material (e.g., books, cell phones, laptops, etc.) is allowed.**
- **Only answers that are given in English will be graded.**

I hereby state that I agree to the declared exam conditions and that I have processed the exam independently only with the approved tools and without outside help.

Student Signature _____

Question	Max. Points	Result
1	13	
2	15	
3	14	
4	15	
5	8	
6	14	
7	8	
8	13	
Sum	100	
Examiner's Signature		

First name	Name	Matr.-Nr.

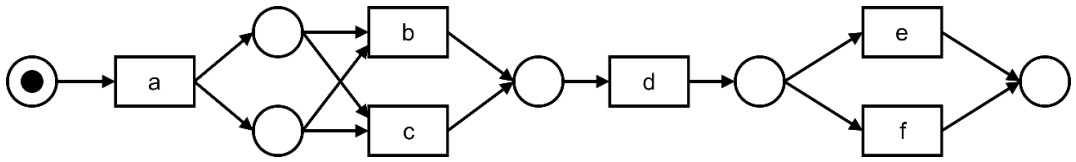
1. Alpha Miner (13 points)

a. Use the Alpha Miner algorithm to construct a Petri net from the following footprint matrix. Show the resulting Petri net. The only start activity is **b**, the only end activity is **e**. **(7 points)**

	a	**b**	**c**	**d**	**e**	**f**
a	#	←	→	→	#	#
b	→	#	#	#	#	#
c	←	#	#	||	→	#
d	←	#	||	#	→	→
e	#	#	←	←	#	←
f	#	#	#	←	→	||

First name	Name	Matr.-Nr.

b. Assume you have an event log $L = [\langle a, b, d, e\rangle^{15}, \langle a, c, d, f\rangle^{16}, \langle a, b, h, e\rangle^{2}]$ and the following Petri net M_b. **(6 points)**



You want to manually change the model to improve its quality measures. For each of the following quality measures, **please modify the Petri net M_b** to improve this measure. For each measure, modify M_b and do not use your modified model from a previous quality measure. Apply as few changes as possible: add/remove only one place/transition and its corresponding arcs.

i. Fitness (Fraction of traces of the event log that are allowed in the model) **(2 points)**

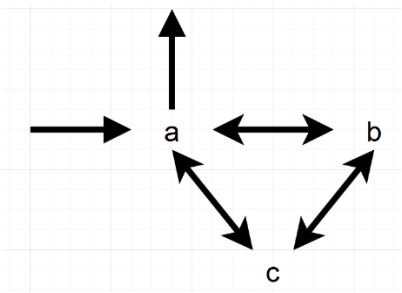
ii. Precision (Fraction of traces allowed by the model that are contained in the event log) **(2 points)**

iii. Simplicity (Number of places and arcs), but don't reduce fitness or precision! **(2 points)**

First name	Name	Matr.-Nr.

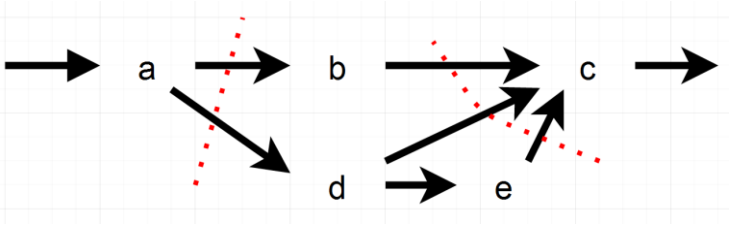
2. Inductive Miner (15 points)

a. Consider the following directly follows graph. Find the first cut, mark it and give its type. **(2 points)**



b. Consider the event log L and the following directly follows graph with the indicated cut. Give the type of the cut and make the log projections for L. **(3 points)**

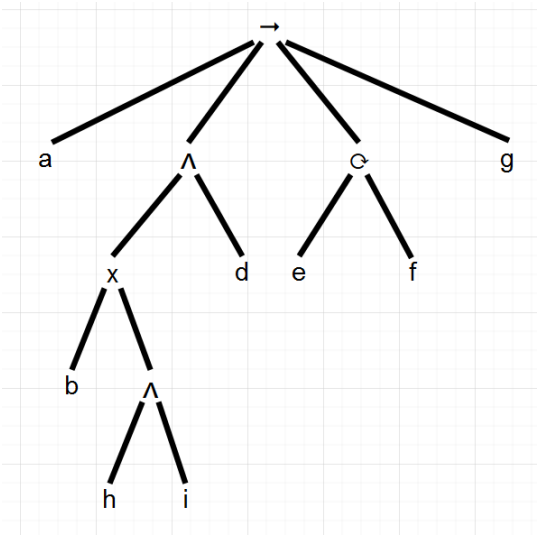
$$L = [\langle a, b, \mathrm{c}\rangle, \langle a, d, \mathrm{e}, \mathrm{c}\rangle, \langle a, d, c\rangle]$$



First name	Name	Matr.-Nr.

c. Give a simple example DFG and event log with a correct sequence cut that has at least one empty trace in one of the log projections. **(2 points)**

d. For the following process tree, give the Petri net with the same behavior. No duplicate transitions are allowed. **(4 points)**

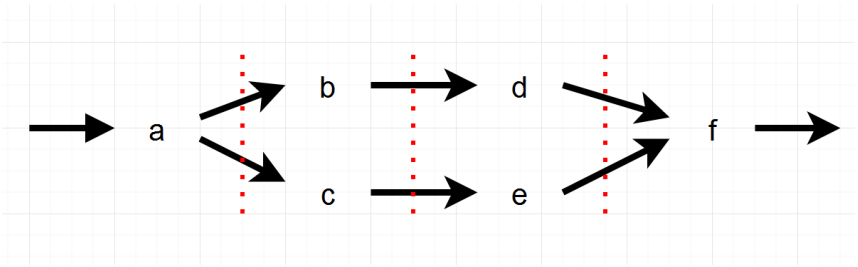


First name	Name	Matr.-Nr.

e. Consider the following directly follows graph and the indicated cut with resulting subsets $A_1 = \{a\},\ A_2 = \{b, c\},\ A_3 = \{d, e\}$ and $A_4 = \{f\}$. **(4 points)**

i. Is this a correct sequence cut w.r.t. the definitions? Shortly explain your answer

ii. If the cut is not fulfilling the definition, what needs to be changed in the directly follows graph (adding or deleting arcs) so that the cut is correct. If the cut fulfills the definition create the DFGs of the next step without log projection, i.e., create the DFGs you would use for finding the next cut.



First name	Name	Matr.-Nr.

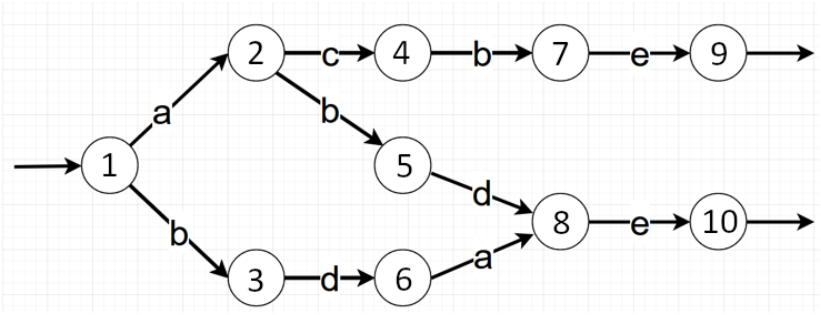
3. Region-based Miner (14 points)

a. Construct the transition system applying the ***past with set abstraction*** for the event log L_1. (**2 points**)

$$L_1 = [\langle a, b, a, c\rangle^5, \langle b, a, a, c\rangle^3]$$

Assume the following transition system based on event log L_2, then answer parts b and c.

$$L_2 = [\langle a, c, b, e\rangle^{16}, \langle a, b, d, e\rangle^{19}, \langle b, d, a, e\rangle^{15}]$$



b. Specify one possible abstraction that leads to the provided transition system. **(1 point)**

c. Give 6 minimal non-trivial regions. Specify only the set of states for each of these regions. Do not use label splitting. You do not have to find all minimal regions. (**6 points**)

First name	Name	Matr.-Nr.

d. Assume the four minimal regions in the following table. The underlying transition system has states {1,2,3,4,5,6,7} and is created by the event log L_3 with **activities {a,c,d,e,f}.**

$$L_3 = [\langle a, c, d, e\rangle^5, \langle a, c, d, f\rangle^3, \langle a, d, c, f\rangle]$$

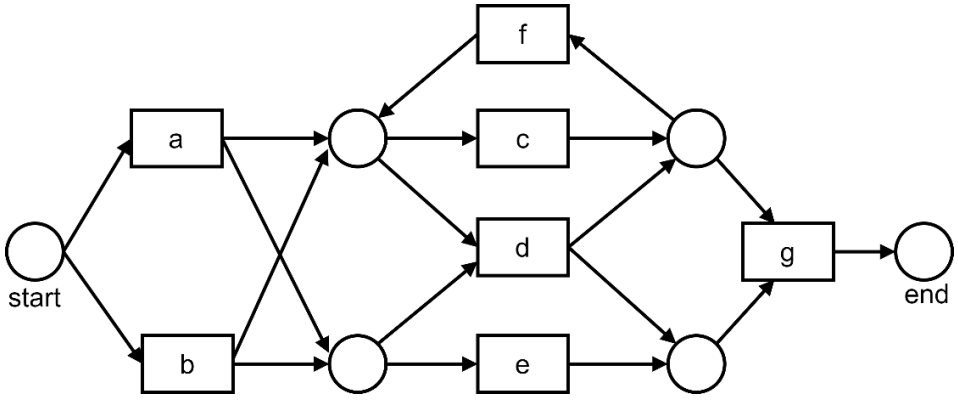
Region	**Incoming**	**Outgoing**
1	/	a
2,3,4,5	a	e,f
6	e	/
7	f	/

Create the Petri net corresponding to the minimal regions. **(5 points)**

First name	Name	Matr.-Nr.

4. Footprint and Token-based Conformance (15 points)

Consider the following process model and answer parts a and b.



a. For the following event log

$$L_1 = [\langle b, a, c, e, f, d, f, c, f, d, g\rangle, \langle a, b, d, f, e, g\rangle]$$

Complete the footprint matrix for the process model and the event log. Then calculate the footprint-based conformance. Please specify the formula that you use. **(7 points)**

	Process Model						
	a	**b**	**c**	**d**	**e**	**f**	**g**
a	#					#	#
b		#				#	#
c			#	#			
d			#	#	#		
e				#	#		
f	#	#				#	
g	#	#					#

	Event Log						
	a	**b**	**c**	**d**	**e**	**f**	**g**
a	#					#	#
b		#				#	#
c			#	#			
d			#	#	#		
e				#	#		
f	#	#				#	
g	#	#					#

First name	Name	Matr.-Nr.

b. For the provided process model and the following event log

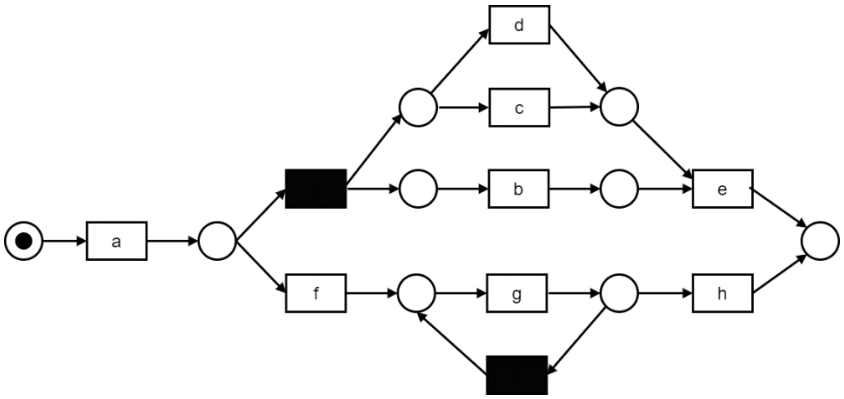
$$L_2 = [\langle a, b, c, g\rangle^2, \langle a, e, d, f, g\rangle^3]$$

Calculate the number of produced, consumed, missing, and remaining tokens for each trace variant. Then calculate the token-based replay fitness for this event log. Please specify the formula that you use. **(8 points)**

First name	Name	Matr.-Nr.

5. Alignment-based Fitness (8 points)

Consider the following process model and event log. Also, consider the same costs for both "move on log", "move on model", and no cost for "silent moves".



$$L = [\langle a, f, b, h\rangle, \langle a, g, g\rangle, \langle a, g, b\rangle]$$

a. Specify one optimal alignment for each trace. **(4.5 points)**

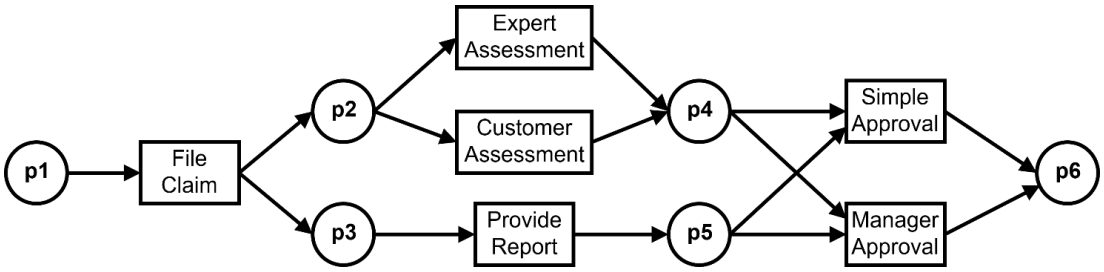
First name	Name	Matr.-Nr.

b. Compute the alignment-based fitness of each trace. Please specify the formula that you use. **(3.5 points)**

First name	Name	Matr.-Nr.

6. Decision Mining (14 points)

Consider the following process model of insurance claim handling in case of water damage and the corresponding event log recorded by an information system.



Case ID	Activity	Transition	Timestamp	Coverage	Cost	Type
1	File Claim	complete	15.12.2020	Extended	20000	Black
1	Expert Assessment	complete	19.12.2020	Extended	20000	Black
1	Provide Report	complete	20.12.2020	Extended	20000	Black
1	Manager Approval	complete	22.01.2021	Extended	20000	Black
2	File Claim	complete	05.01.2021	Standard	2400	Black
2	Customer Assessment	complete	06.01.2021	Standard	2400	Black
2	Provide Report	complete	10.01.2021	Standard	2400	Black
2	Simple Approval	complete	21.01.2021	Standard	2400	Black
3	File Claim	complete	06.01.2021	Extended	35000	Black
3	Provide Report	complete	08.01.2021	Extended	35000	Black
3	Expert Assessment	complete	09.01.2021	Extended	35000	Black
3	Manager Approval	complete	25.01.2021	Extended	35000	Black
4	File Claim	complete	07.01.2021	Standard	1900	Black
4	Provide Report	complete	08.01.2021	Standard	1900	Black
4	Customer Assessment	complete	12.01.2021	Standard	1900	Black
4	Simple Approval	complete	18.01.2021	Standard	1900	Black
5	File Claim	complete	12.01.2021	Extended	12000	Clean
5	Provide Report	complete	16.01.2021	Extended	12000	Clean
5	Expert Assessment	complete	17.01.2021	Extended	12000	Clean
5	Simple Approval	complete	28.01.2021	Extended	12000	Clean
6	File Claim	complete	14.01.2021	Standard	15000	Black
6	Provide Report	complete	15.01.2021	Standard	15000	Black
6	Customer Assessment	complete	16.01.2021	Standard	15000	Black
6	Manager Approval	complete	02.02.2021	Standard	15000	Black

First name	Name	Matr.-Nr.

a. Consider place p4 as a decision point. Assuming that the damage costs are grouped into *High* (> 10000) and *Low* (≤ 10000), create a table for the corresponding classification problem by showing the values of the descriptive features and the target feature for each case. **(2 points)**

b. Using the table from **a.**), discover a decision tree for this decision point (p4). Choose the data attributes so that information gain is maximized at each step. You do not need to show the computation of the information gain. Add the discovered guards to the model. **(7 points)**

First name	Name	Matr.-Nr.

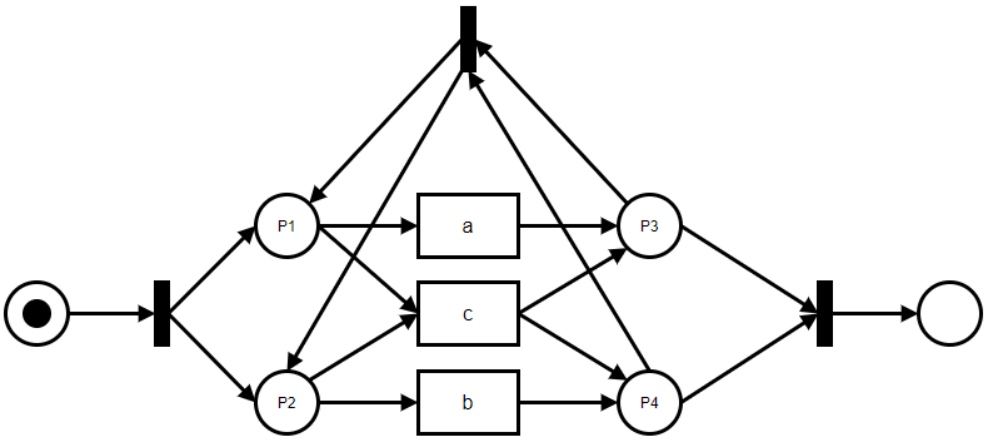
c. Suppose that a new insurance customer with an *Extended* coverage had a sudden leak of his washing machine causing a damage of type *Clean* that costs 12000 euros. Give an expected sequence of activities for this customer and explain your choice for each decision point in the model. **(5 points)**

First name	Name	Matr.-Nr.

7. Performance analysis (8 points)

Consider the following event log and process model.

Case ID	Activity	Lifecycle Transition	Timestamp
1	a	start	0
1	a	complete	6
1	b	start	2
1	b	complete	3
1	c	start	10
1	c	complete	14
1	a	start	17
1	a	complete	19
1	b	start	16
1	b	complete	23
2	a	start	3
2	a	complete	10
2	b	start	1
2	b	complete	9
2	c	start	15
2	c	complete	18
2	c	start	20
2	c	complete	26



First name	Name	Matr.-Nr.

a. Assume that the invisible transitions in the model fire as soon as they are enabled, and require time 0 to fire (they fire instantaneously). Compute the **average waiting times** of the places P1 through P4 for each case in the event log. **(4 points)**

b. In this process model, which two places can be considered as bottlenecks in terms of **average waiting time** across all cases? Explain why the waiting time is shorter in the other two places? **(3 points)**

c. In terms of **total waiting time across all of the four places**, which case is the most performant (has the shortest waiting time)? **(1 point)**

First name	Name	Matr.-Nr.

8. Organizational Mining (13 points)

a. Consider the following event log.

$$L = [\langle a, b, c, d, e\rangle, \langle a, c, b, d, f\rangle, \langle a, d, b, b, e\rangle, \langle a, c, d, b, a\rangle]$$

Compute the **dependency matrix**. Provide the formula you use for the dependency measure. **(8 points)**

First name	Name	Matr.-Nr.

b. Consider the event log made of the following traces, given as pairs of sequences of activities (a through e) and resources carrying out each activity in order (U1 through U4):

$$L = [\langle a^{U1}, b^{U2}, c^{U1}, d^{U1}, e^{U4}\rangle, \langle a^{U2}, c^{U3}, b^{U1}, d^{U3}, e^{U2}\rangle, \langle a^{U1}, d^{U1}, b^{U3}, c^{U2}, e^{U4}\rangle, \langle a^{U4}, b^{U1}, d^{U2}, c^{U2}, e^{U4}\rangle]$$

Furthermore, consider the associated dependency matrix:

	a	**b**	**c**	**d**	**e**
a	0.00	0.67	0.50	0.50	0.00
b	-0.67	0.00	0.25	0.25	0.00
c	-0.50	-0.25	0.00	0.00	0.67
d	-0.50	-0.25	0.00	0.00	0.67
e	0.00	0.00	-0.67	-0.67	0.00

Compute the real handover of work graph, considering a dependency threshold of **0.3** and a handover of work threshold of **0.4**. Do not consider multiple occurrences of handovers in the same trace. **(5 points)**

First name	Name	Matr.-Nr.

First name	Name	Matr.-Nr.

Additional pages for calculation

First name	Name	Matr.-Nr.