

Wie laufen die realen Prozesse in meiner Organisation?

Ein kleiner Einblick in Process Mining

Tobias Seyffarth
JCNetwork Alumni

Vorstellung und Eure Erwartungen an den Workshop.

- Wie heiße ich?
- Was studiere ich?
- In welchen Verein bin ich tätig?
- Was sind meine Erwartungen an den Workshop?
- Welche Berührungspunkte zum hatte ich bereits zum Geschäftsprozessmanagement?



Lernziele des Workshops.

- Sie verstehen, welche Informationssysteme die notwendigen Log-Dateien für das Process Mining zur Verfügung stellen können.
- Sie können ausgewählter Methoden des Process Mining anwenden und kennen ihre jeweiligen Möglichkeiten und Grenzen.
- Sie kennen verschiedene Anwendungsszenarien für das Process Mining.
- Sie kennen verwandte Anwendungsfelder des Process Mining.



Inhalte des Workshops

1

Wozu Process Mining und mit welcher Datenbasis?

2

Heuristic Miner

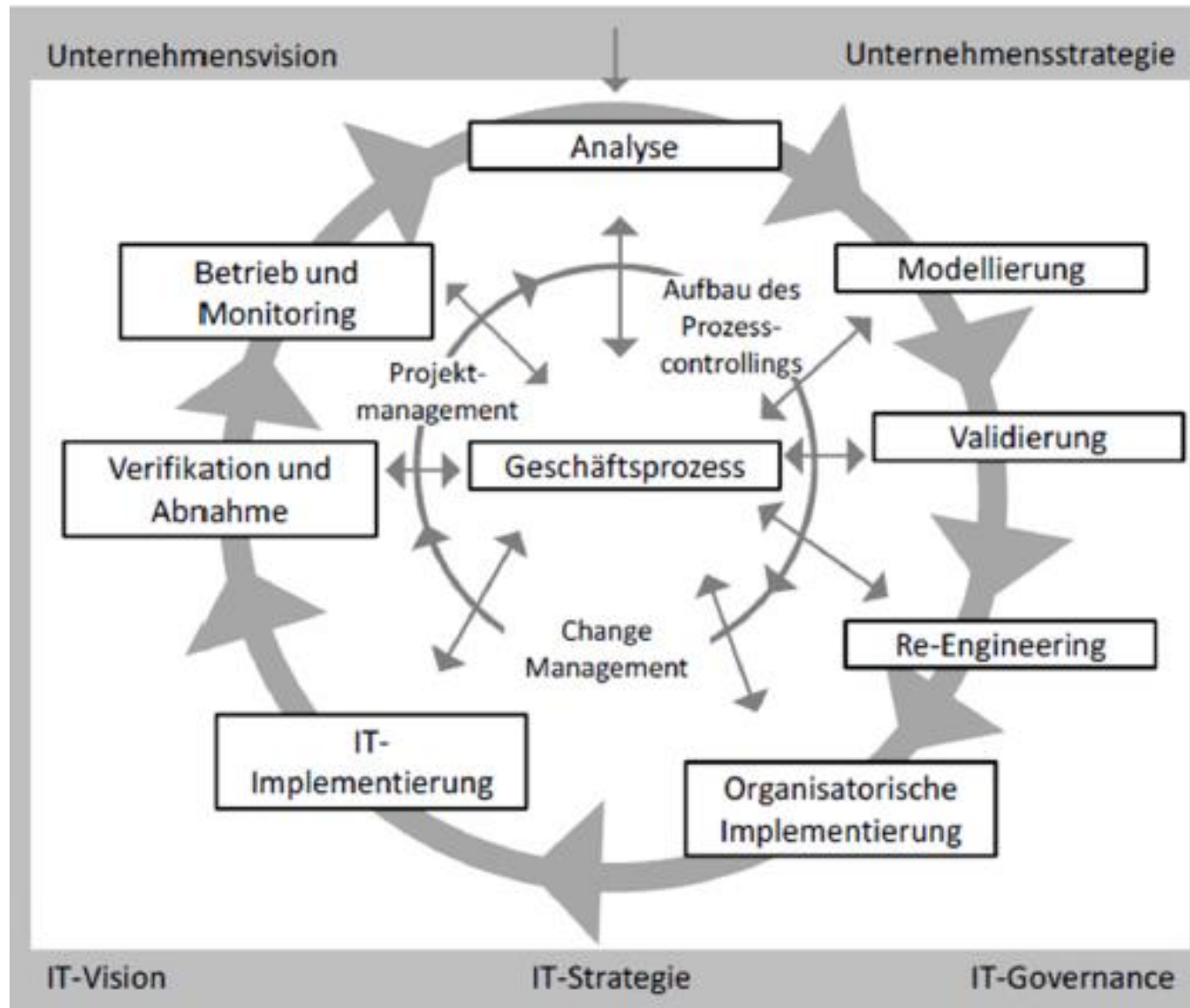
3

Inductive Miner

4

Weitere Themenfelder im Process Mining

GPM ist keine einmalige Sache und Bedarf als Startpunkt einer Modellierung von Prozessen.



Modellierung

Zusammenfassen der Prozesse zu einem Modell in unterschiedlichen Formen.

Analyse

Analyse auf dem Prozessmodell

Analyse der real ausgeführten Prozesse

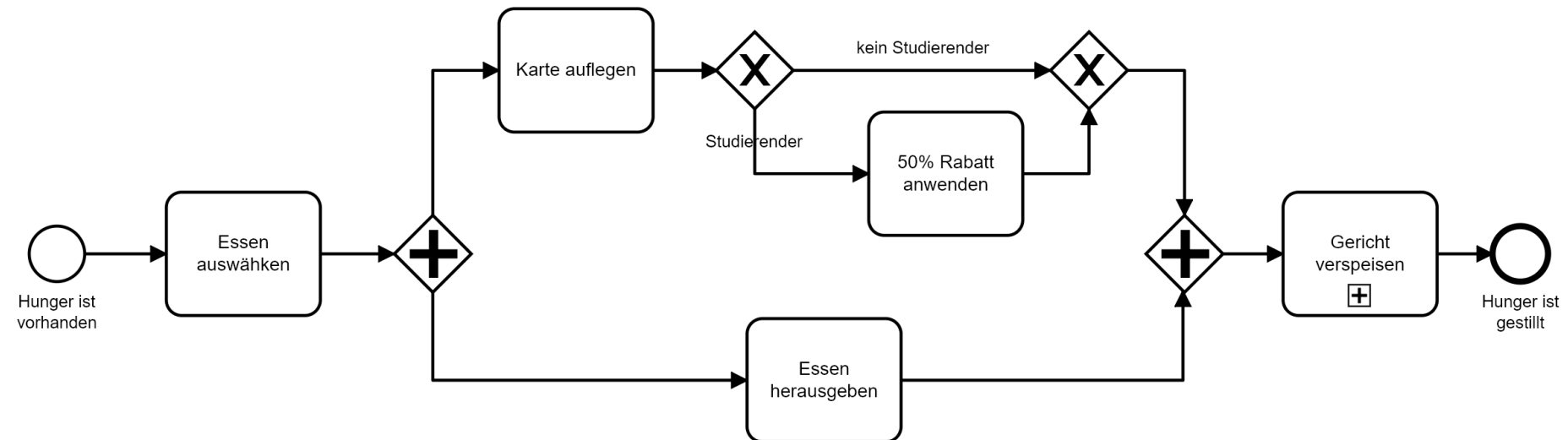
Funk (2010), S. 15.

Die Modellierung von Geschäftsprozessen kann auf verschiedene Arten erfolgen.

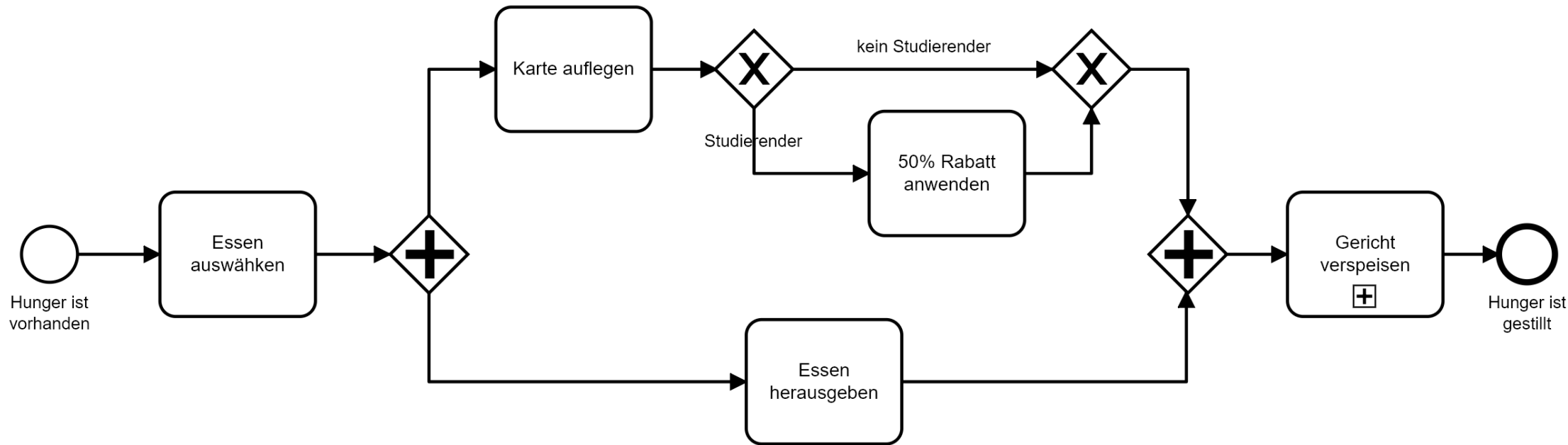
Textuelle
Beschreibung

- Sollte Hunger vorhanden sein, wählt der Kunde sein Essen an der Ausgabe aus.
- Zur Bezahlung legt er seine Karte auf das Lesegerät. Sollte er ein Studierender sein, wird auf den Essenspreis ein Rabatt von 50% gewährt.
- Parallel zur Bezahlung wird sein Essen herausgegeben.
- Im Anschluss verspeist der Kunde seine Mahlzeit womit sein Hunger gestillt ist.

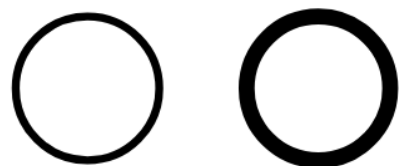
BPMN-
Modellierung



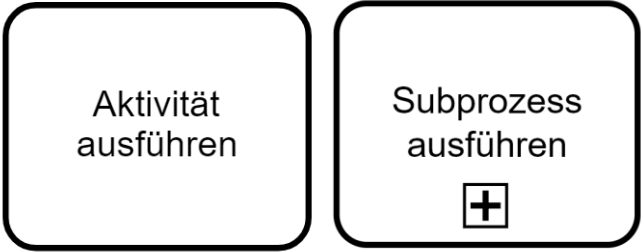
Geschäftsprozesse können bereits mit wenigen Elementen der BPMN modelliert werden.



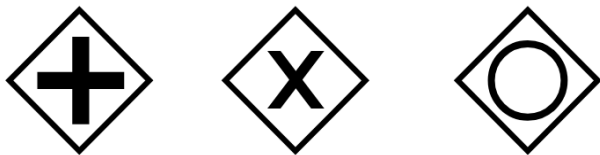
Ereignisse



Aktivitäten & Subprozesse

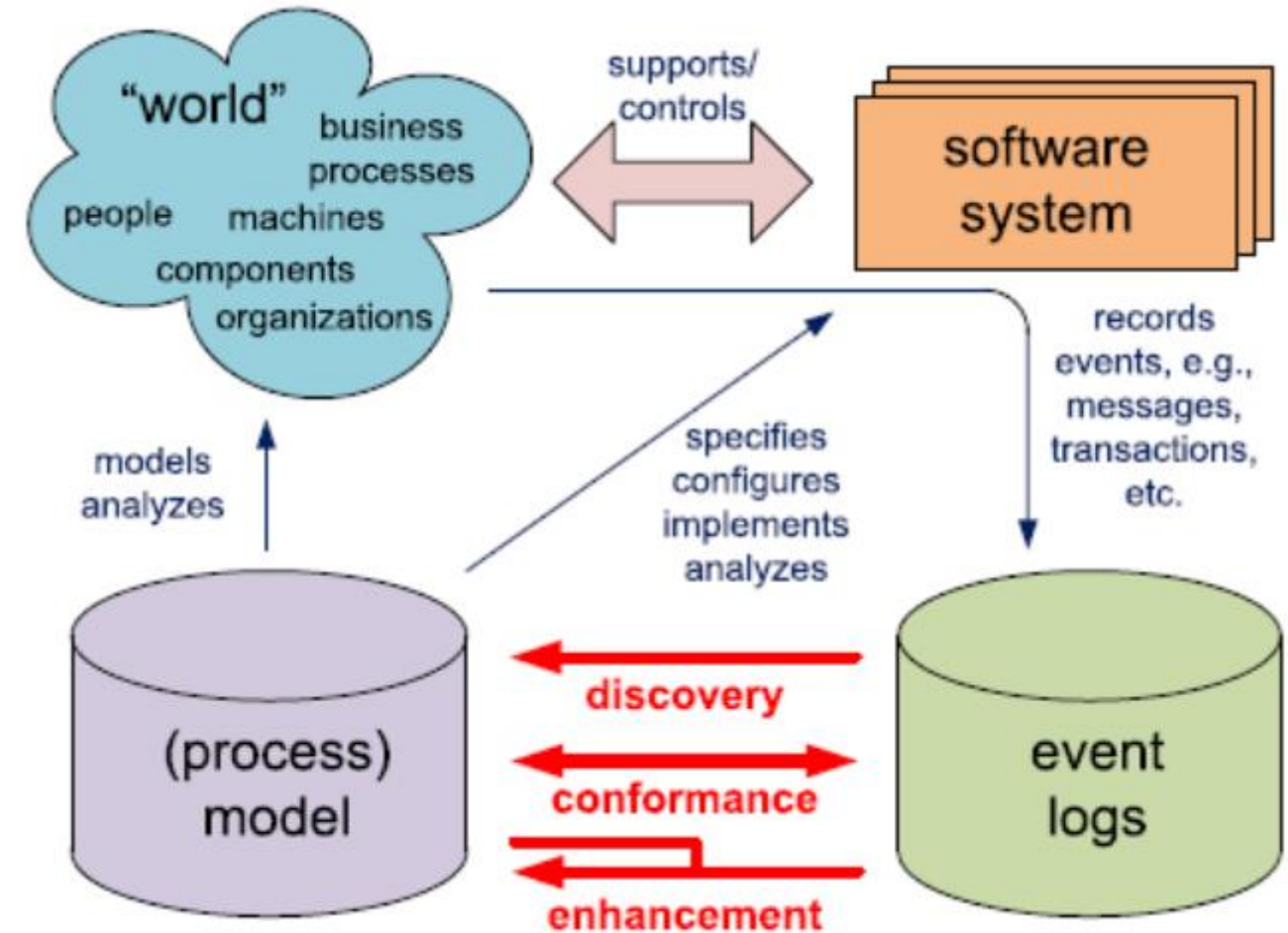


Gateway



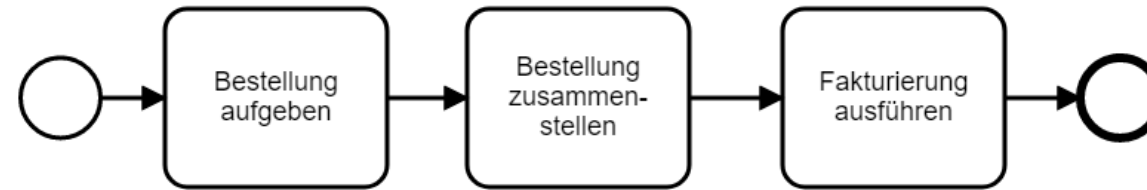
Ein Definitionsversuch des Process Mining.

- Process Mining ist eine Forschungsdisziplin an der Schnittstelle zwischen Data Mining und Machine Learning auf der einen und Prozessmodellierung sowie –analyse auf der anderen Seite.
- Die Idee von Process Mining ist das “Entdecken”, Überwachen und Verbessern von realen (und nicht angeommenen) Prozessen durch die Extraktion von Wissen aus Event-Logs, die in IT-Systemen vorliegen.



van der Aalst (2016), S. 31 f.

Event-Logs dienen als Datenbasis für das Process Mining.



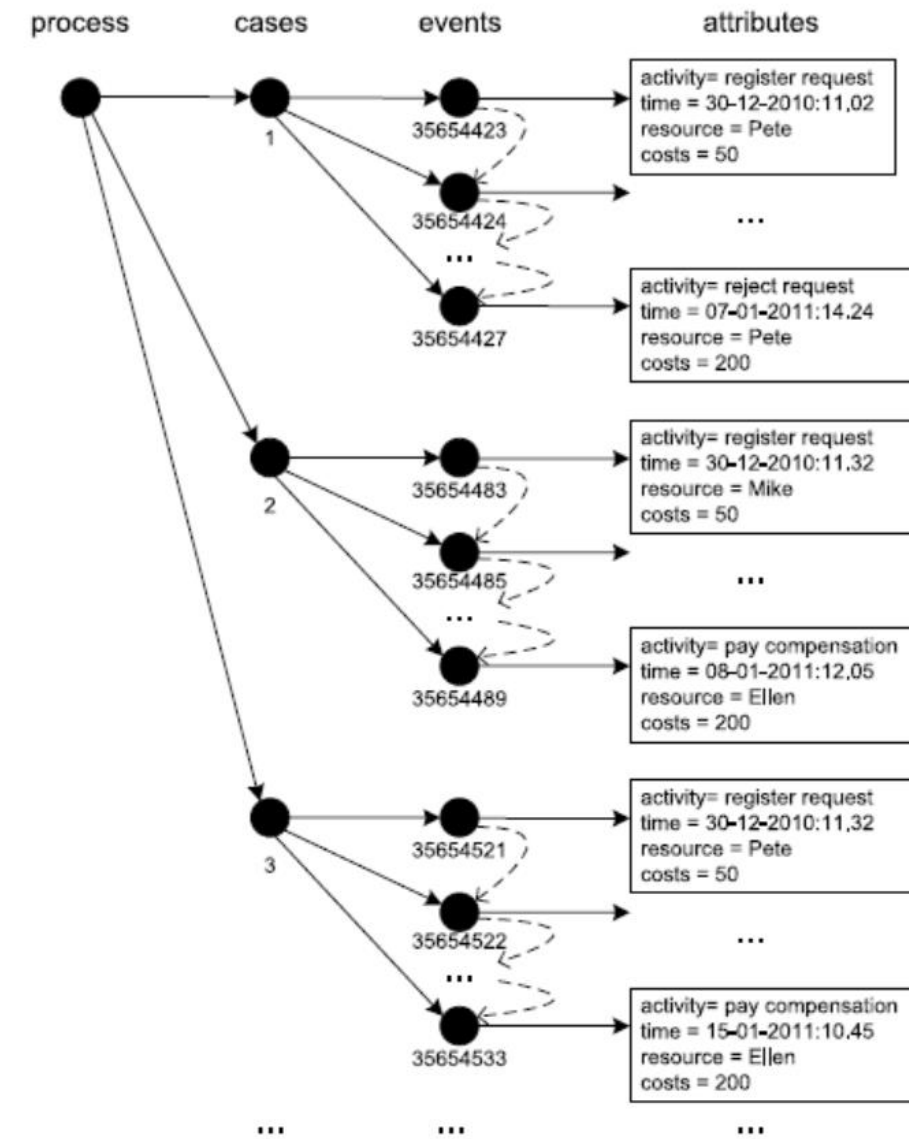
Bestellnummer	Aktivität	Zeitstempel	Ressource	Produkt	Menge	Preis	...
87312	Bestellung aufgeben	2021-05-22@9:15	Max	iPad	2	...	
87313	Bestellung aufgeben	2021-05-22@10:50	Lisa	iPhone	1	...	
87312	Bestellung zusammenstellen	2021-05-22@10:15	Caro	iPad	2		
87312	Fakturierung ausführen	2021-05-22@10:20	Peter	iPad	2		
87314	Bestellung aufgeben		Hilde	iPad	1		
87313	Bestellung zusammenstellen	2021-05-22@10:57	Caro	iPhone	1		
87315	Bestellung aufgeben	2021-05-22@11:20	Susanne	iPad	1		
87313	Fakturierung ausführen	2021-05-22@11:50	Peter	iPhone	1		
...							

[<Bestellung aufgeben, Bestellung zusammenstellen, Fakturierung ausführen>,
<Bestellung aufgeben, Bestellung zusammenstellen, Fakturierung ausführen>,
<Bestellung aufgaben, Bestellung zusammenstellen, ...>]

van der Aalst (2016), S. 129.

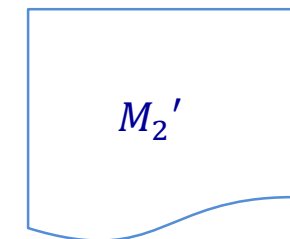
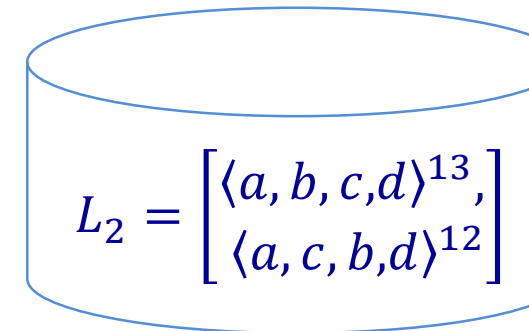
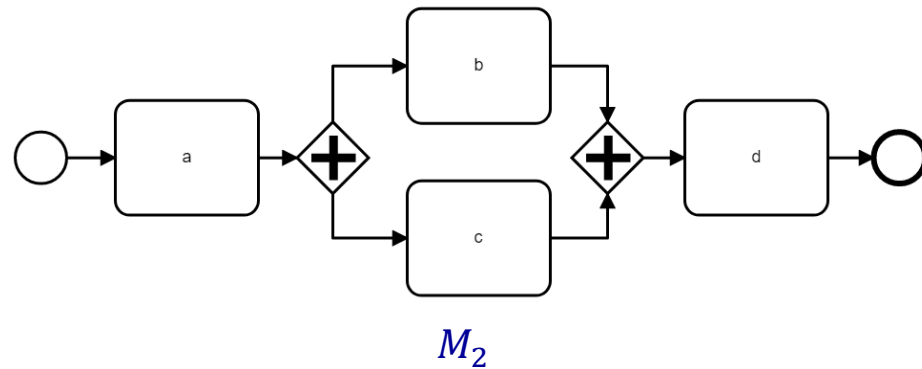
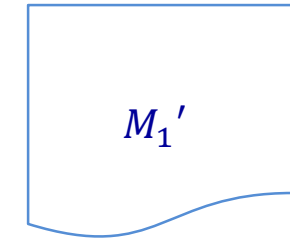
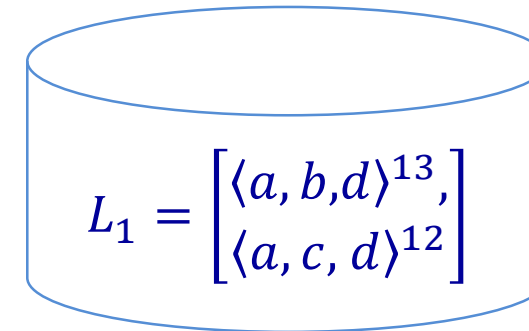
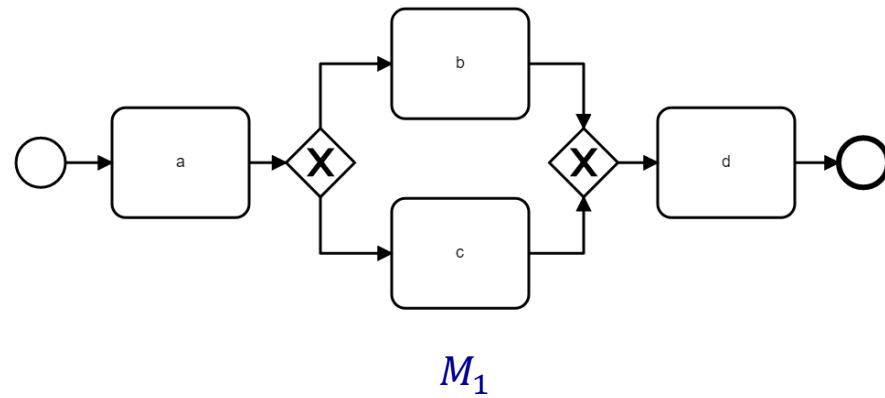
Das Event-Log kann auf wesentliche Informationen zusammengefasst werden.

- Ein Prozess besteht aus Cases (bspw. Bestellnummer).
- Ein Case besteht aus Events, sodass jedes Event zu genau einem Case gehört.
- Events sind innerhalb eines Cases in ihrer zeitlichen Reihenfolge sortiert.
- Events haben Attribute, wie bspw. einen Namen.
- Die geordneten Events in einem Case werden Trace genannt.
- Ein Event-Log ist eine Menge an Traces, wobei ein Trace mehrfach vorkommen kann.
- Ein Trace ist eine Sequenz von Aktivitätsnamen. Wir abstrahieren alle anderen Attribute.
- $L_1 = [\langle a, b, c, d \rangle^3, \langle a, c, b, d \rangle^2, \langle a, e, d \rangle]$



van der Aalst (2016), S. 129 f.

Zusammenfassung: Vom Prozessmodell über das Event-Log zum „entdeckten“ Prozessmodell



van der Aalst (2016), S. 179
<https://bimp.cs.ut.ee/simulator>

Inhalte des Workshops

1

Wozu Process Mining und mit welcher Datenbasis?

2

Heuristic Miner

3

Inductive Miner

4

Weitere Themenfelder im Process Mining

Der Heuristic Miner besteht aus zwei Schritten.

- Schritt 1: Berechnen des Dependency Graphs
- Schritt 2: Erkennen von Splits und Joins

Berechnen des Dependency Graphs: Erstellen der Foot-Print-Matrix

$$L = [\langle a, e \rangle^5, \langle a, b, c, e \rangle^{10}, \langle a, c, b, e \rangle^{10}, \langle a, b, e \rangle^1, \langle a, c, e \rangle^1, \langle a, d, e \rangle^{10}, \langle a, d, d, e \rangle^2, \langle a, d, d, d, e \rangle^1]$$

> L	a	b	c	d	e
a	0	11	11	13	5
b	0	0	10	0	11
c	0	10	0	0	11
d	0	0	0	4	13
e	0	0	0	0	0

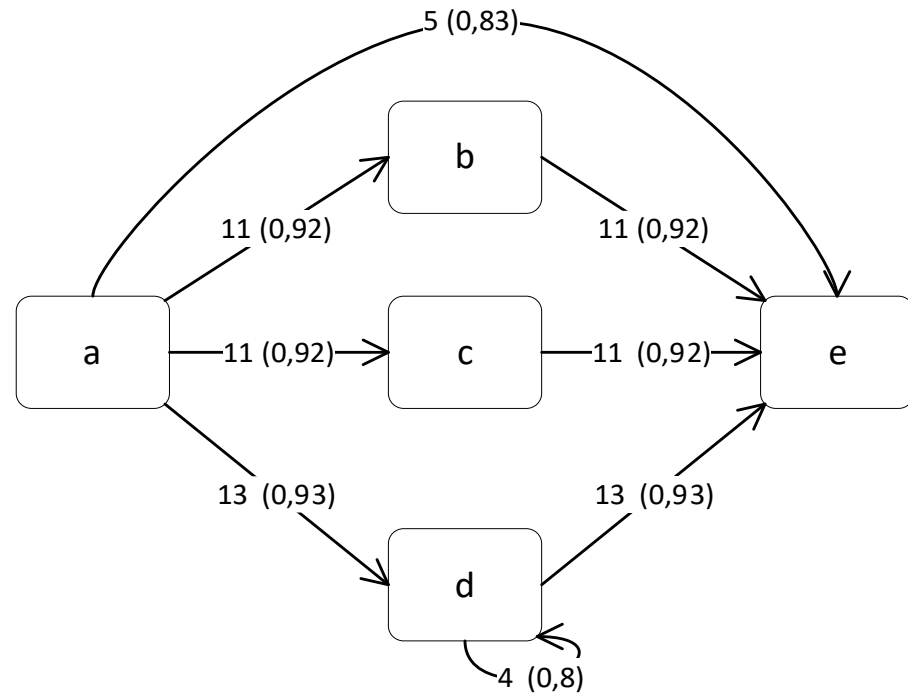
Berechnen des Dependency Graphs: Berechnen der Abhängigkeitsmaße zwischen Events.

> L	a	b	c	d	e
a	0	11	11	13	5
b	0	0	10	0	11
c	0	10	0	0	11
d	0	0	0	4	13
e	0	0	0	0	0

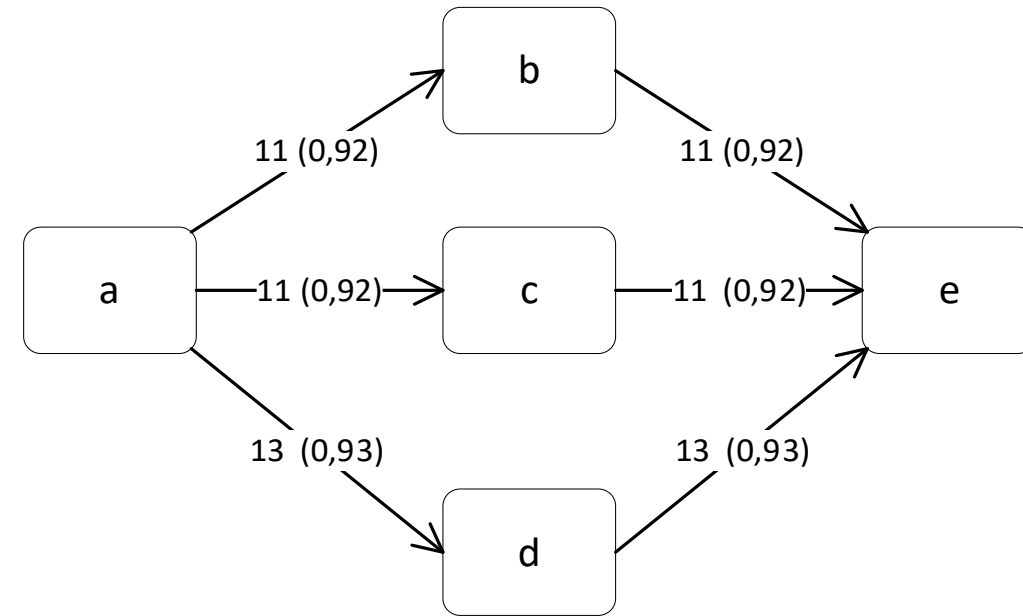
$$|a \Rightarrow b| = \begin{cases} \frac{|a>b| - |b>a|}{|a>b| + |b>a| + 1} & \text{if } a \neq b \\ \frac{a>a}{|a>a| + 1} & \text{if } a = b \end{cases}$$

> L	a	b	c	d	e
a	$\frac{0}{0+1} = 0$	$\frac{11-0}{11+0+1} = 0,92$	$\frac{11-0}{0+11+1} = 0,92$	$\frac{13-0}{0+13+1} = 0,93$	$\frac{5-0}{5+0+1} = 0,83$
b	$\frac{0-11}{0+11+1} = -0,92$	$\frac{0}{0+1} = 0$	$\frac{10-10}{10+10+1} = 0$	$\frac{0}{0+1} = 0$	$\frac{11-0}{0+11+1} = 0,92$
c	$\frac{0-11}{0+11+1} = -0,92$	$\frac{10-10}{10+10+1} = 0$	$\frac{0}{0+1} = 0$	$\frac{0}{0+0+1} = 0$	$\frac{11-0}{0+11+1} = 0,92$
d	$\frac{0-13}{0+13+1} = -0,93$	$\frac{0}{0+1} = 0$	$\frac{0}{0+0+1} = 0$	$\frac{4}{4+1} = 0,8$	
e	$\frac{-5}{0+5+1} = -0,83$	$\frac{-11}{0+11+1} = -0,92$	$\frac{-11}{0+11+1} = -0,92$	$\frac{-13}{0+13+1} = -0,93$	$\frac{0}{0+1} = 0$

Durch die Veränderung der Thresholds kann sich der Dependency Graph verändern.



$$|a \Rightarrow b| > 0,7$$

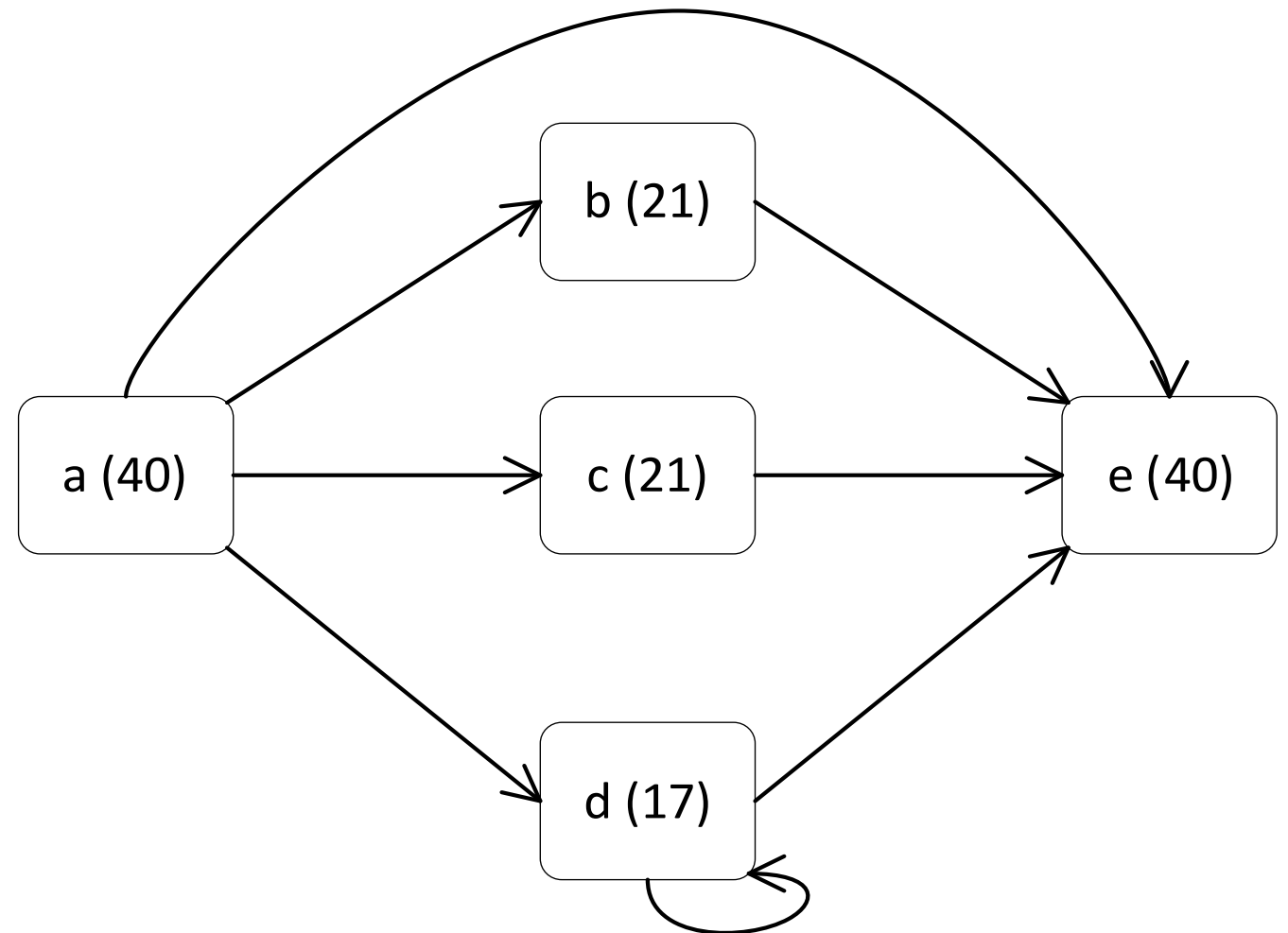


$$|a \Rightarrow b| > 0,9$$

Zusammenfassung zur Berechnung des Dependency Graphs

- Bestimme die Anzahl der direkten Nachfolger in einer Footprint-Matrix.
- Zähle die direkten Nachfolger.
- Berechne die Abhängigkeitsmaße zwischen den Events.
- Zeichne den Dependency Graph, wobei nur Kanten berücksichtigt werden, die die Tresholds aufweisen

$ \triangleright L $	a	b	c	d	e
a	0	11	11	13	5
b	0	0	10	0	11
c	0	10	0	0	11
d	0	0	0	4	13
e	0	0	0	0	0



Hinweise zu den Übungsaufgaben.

- Bitte bilden Sie 2-3 Gruppen.
- Zu jeder der beiden Übungsaufgaben gibt es die Fallbeschreibung A und die Fallbeschreibung B.
- Nach der Bearbeitung der jeweiligen Übungsaufgabe stellen zwei Gruppen ihre Lösungsvorschlag dem Auditorium vor.
- Die Idee des Heuristic Miners lernt man nur durch modellieren und diskutieren. Nutzen Sie die Möglichkeit zur Diskussion, auch wenn Ihnen Ihre Lösung nicht perfekt erscheint!
- Hinweis zur Bearbeitung:
 - Lesen Sie zunächst die komplette Aufgabe.
 - Skizzieren Sie Ihre Lösung auf einem A4-Blatt.
 - Übertragen Sie Ihre Lösung in eine PowerPoint-Folie.



Übungsaufgabe 1 (Anwenden des Heuristic Miner)

Gruppe A

$$L_A = \begin{bmatrix} \langle a, b \rangle^{20}, \\ \langle a, c, d, e \rangle^{10}, \\ \langle a, c, e, d \rangle^{10} \end{bmatrix}$$

Gruppe B

$$L_B = \begin{bmatrix} \langle a, b, c \rangle^{20}, \\ \langle a, c, b \rangle^{20}, \\ \langle a, d \rangle^{10}, \\ \langle a, e \rangle^{10} \end{bmatrix}$$

- Erzeugen Sie aus dem gegebenen Prozess-Log ein Casual-Net.
- Zusatz: Bilden Sie ihr Casual Net in einem BPMN-Prozessmodell ab.
Diskutieren Sie mögliche Stärken und Schwächen Ihres Modells im Hinblick auf das zugrundeliegende Event-Log.



Break: Let's Talk about Podcast



Die Lage der Nation

<https://www.kuechenstud.io/lagedernation/>



Logbuch Netzpolitik

<https://logbuch-netzpolitik.de/>



Forschergeist

<https://forschergeist.de>

Inhalte des Workshops

1

Wozu Process Mining und mit welcher Datenbasis?

2

Heuristic Miner

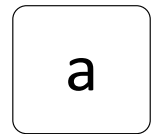
3

Inductive Miner

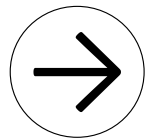
4

Weitere Themenfelder im Process Mining

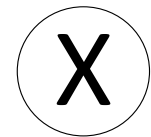
Der Inductive Miner nutzt zur Darstellung den Process Tree.



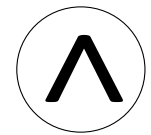
Aktivität



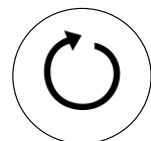
Sequenz



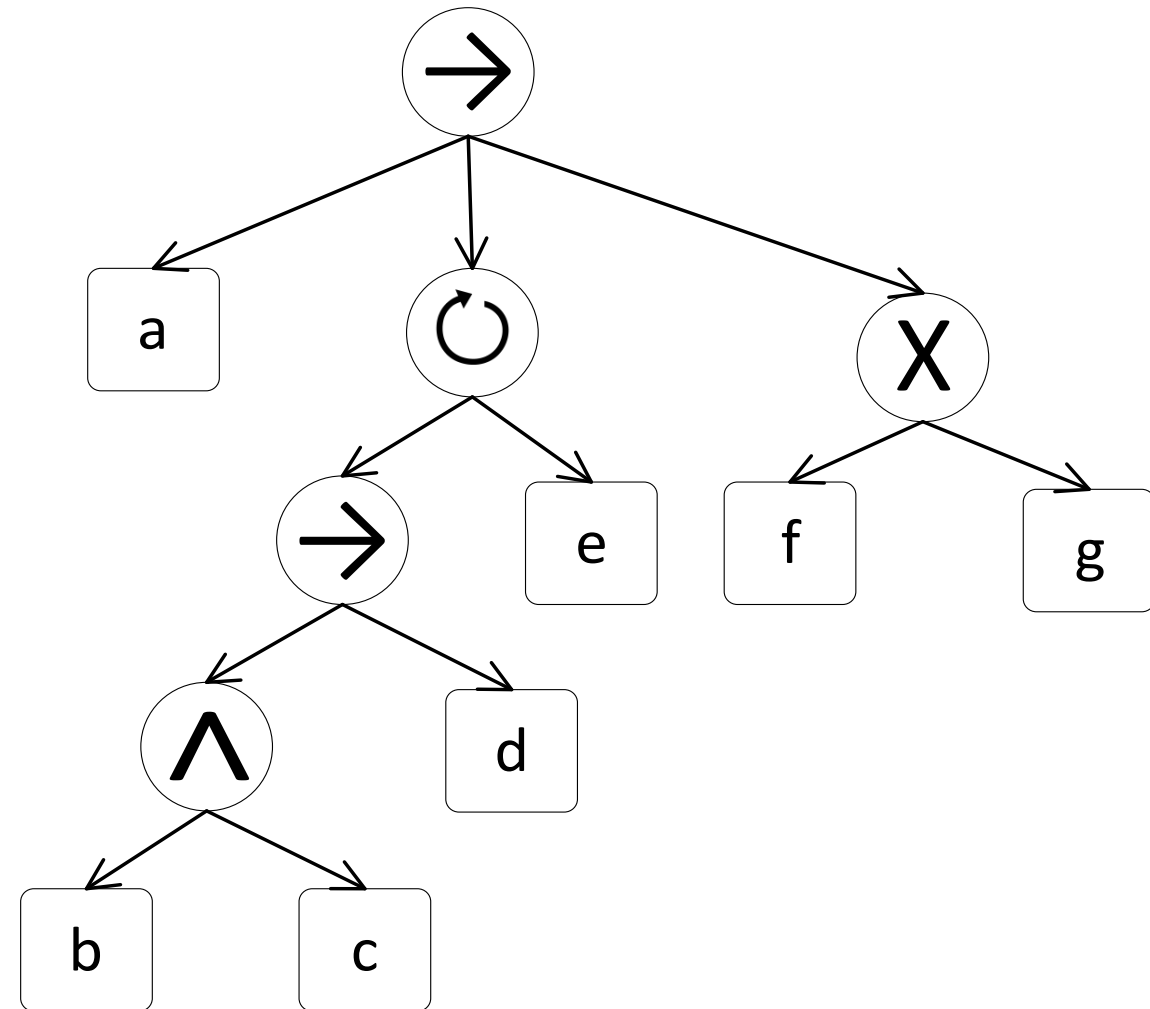
Exklusives Oder



Parallelität

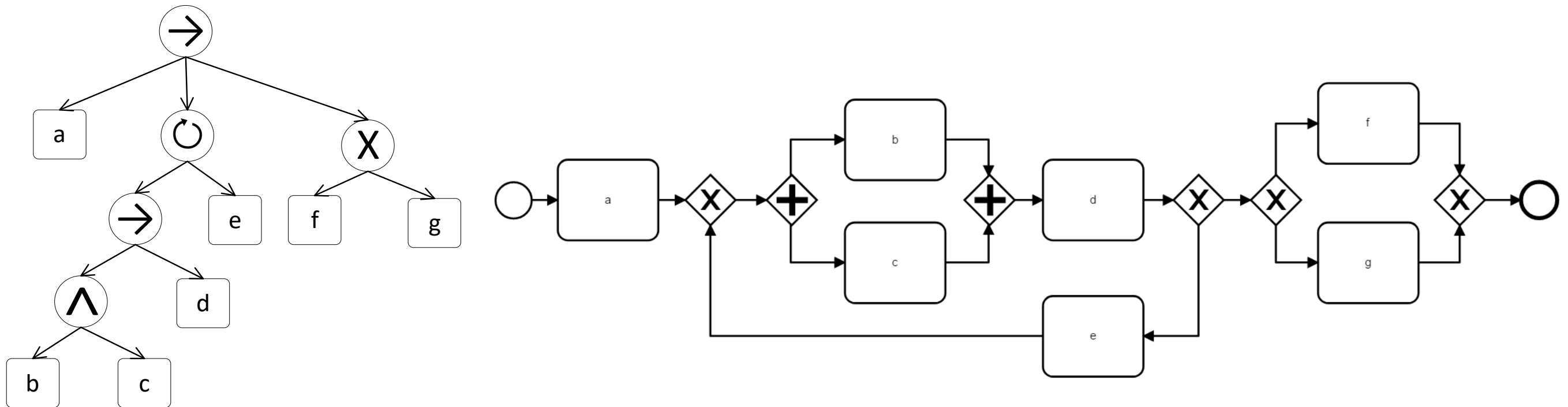


Schleife

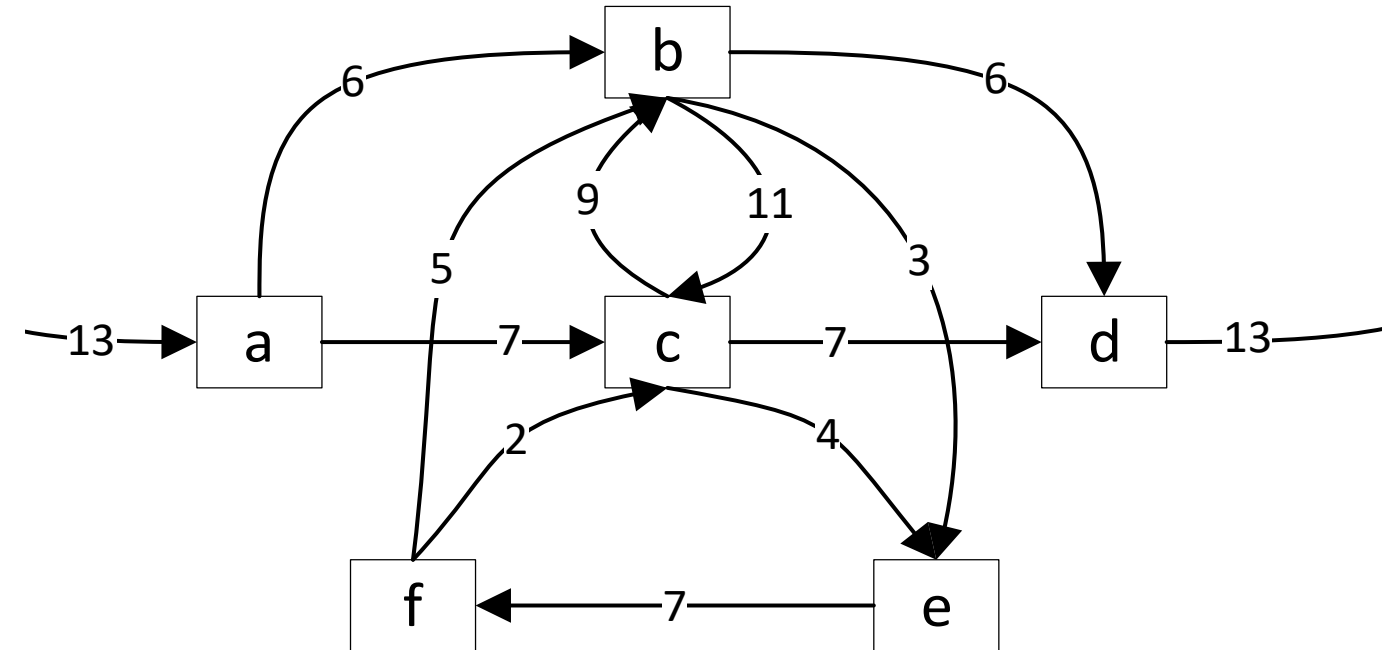
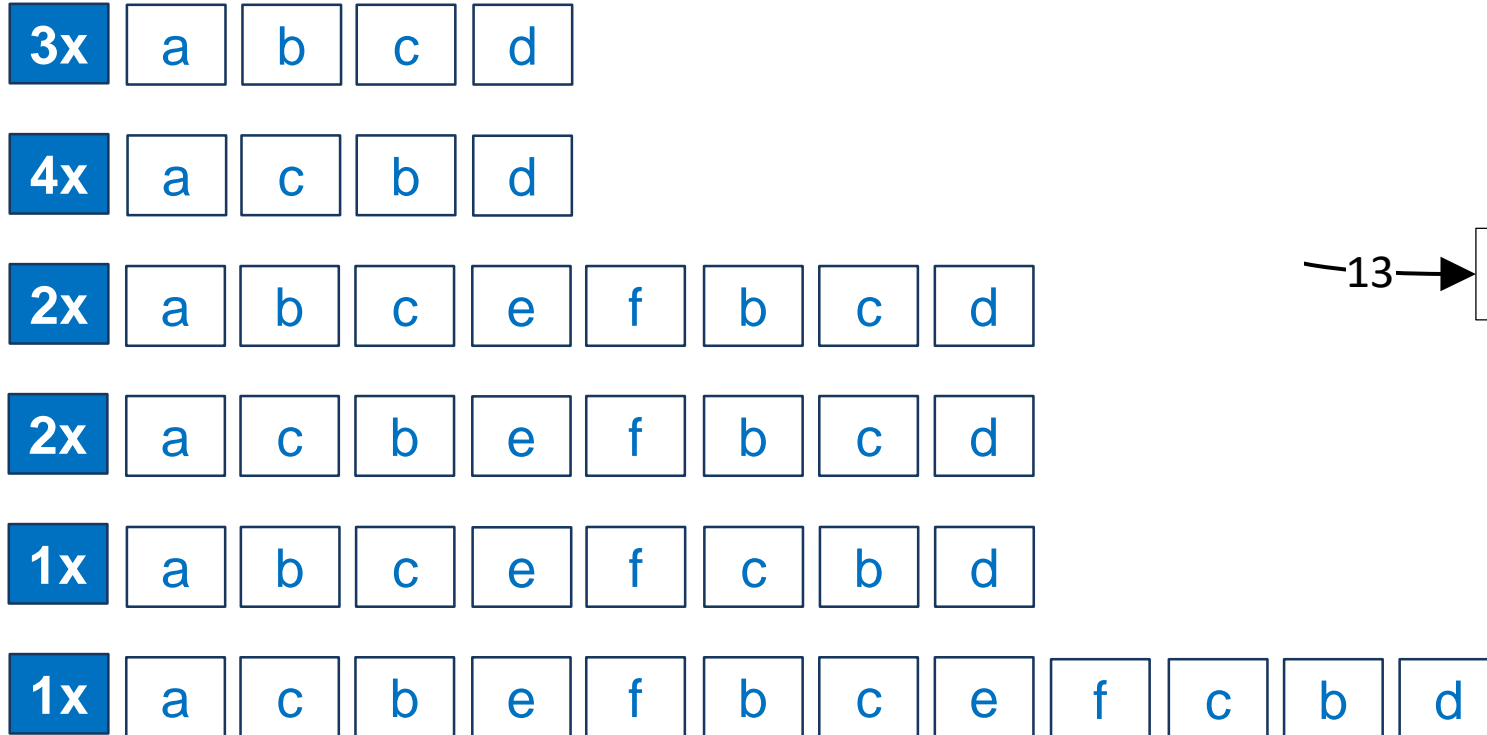


van der Aalst (2016), S. 78 f.

Der Inductive Miner nutzt zur Darstellung den Process Tree.

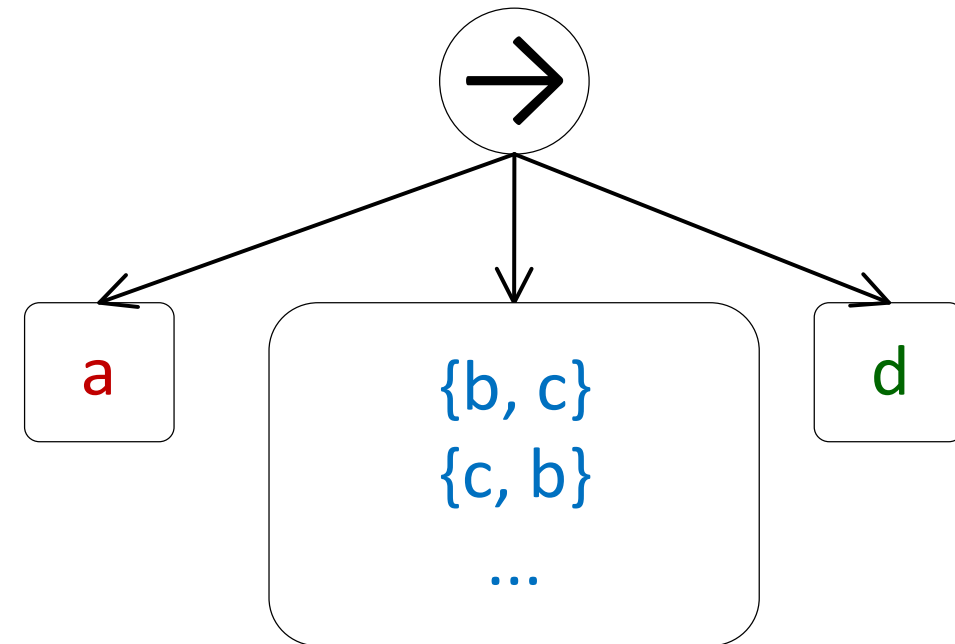
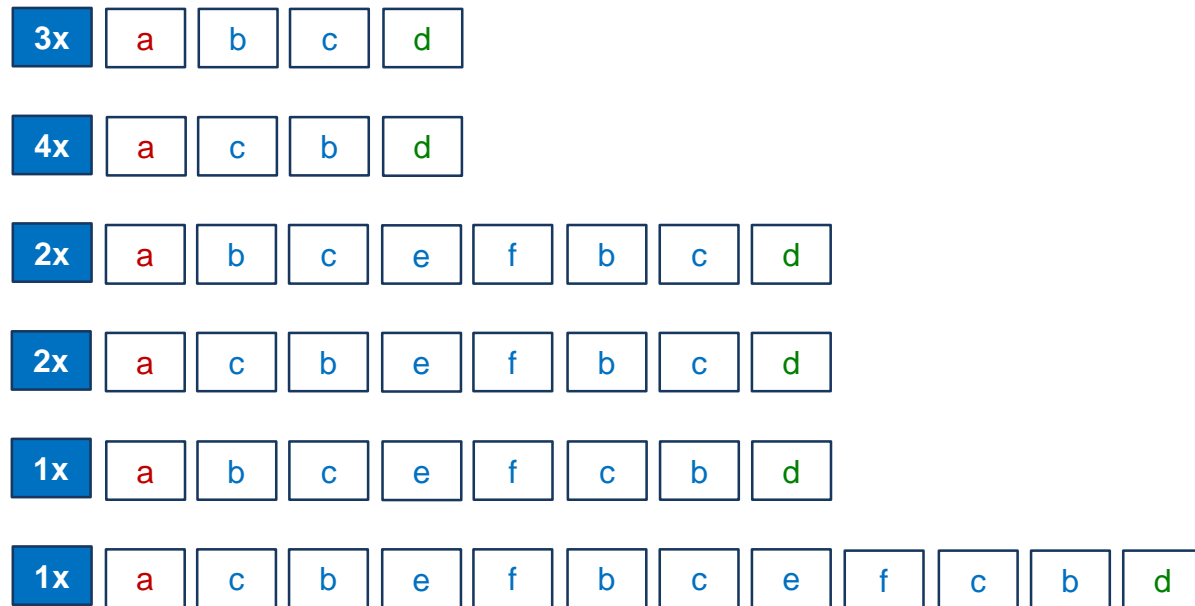
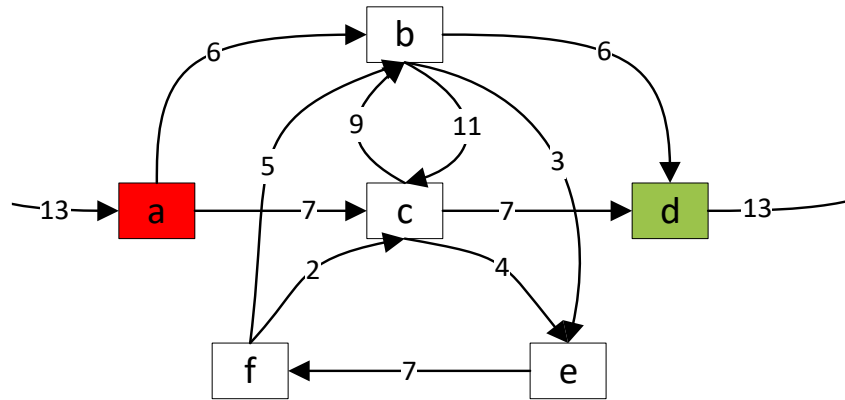


Zunächst wird ein Directly-Follows-Graph erstellt.



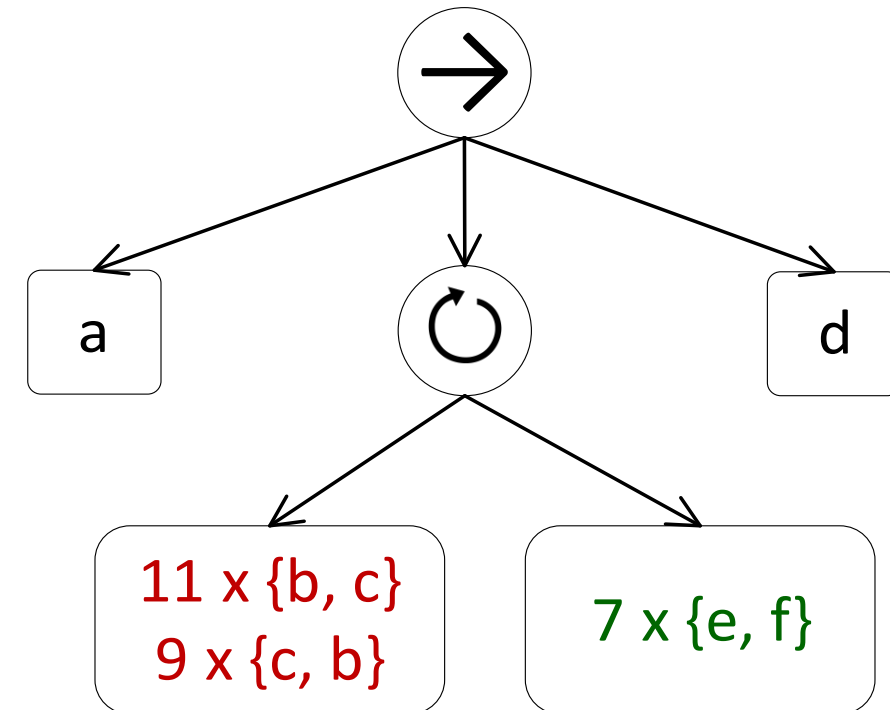
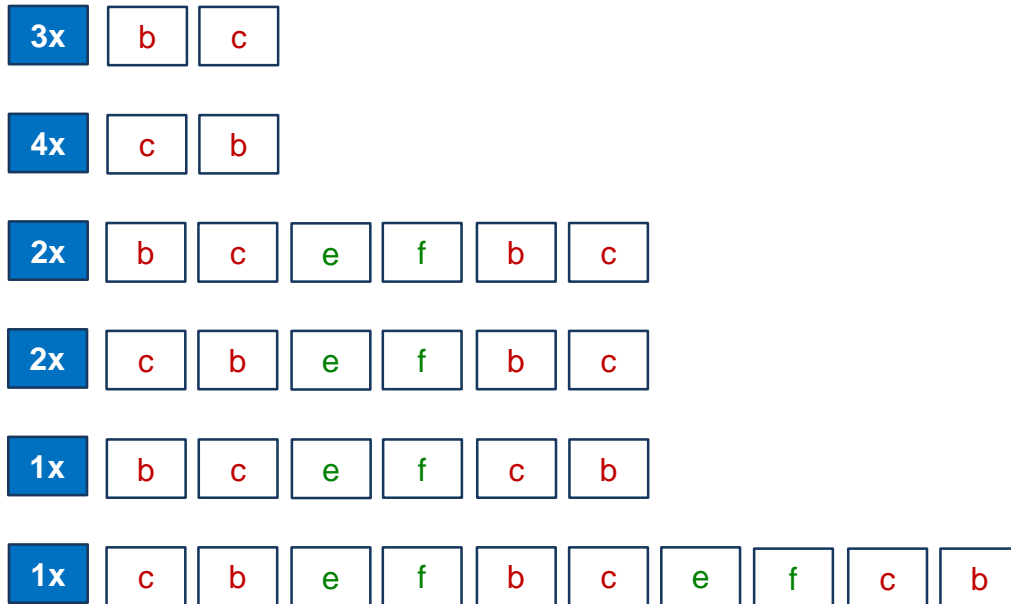
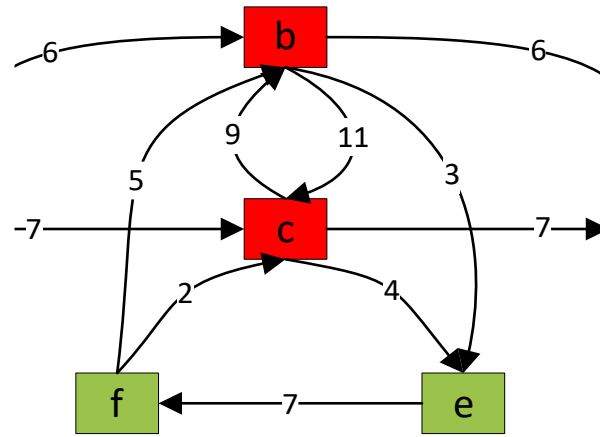
van der Aalst (2016), S. 222 ff.

Sequence Cut.



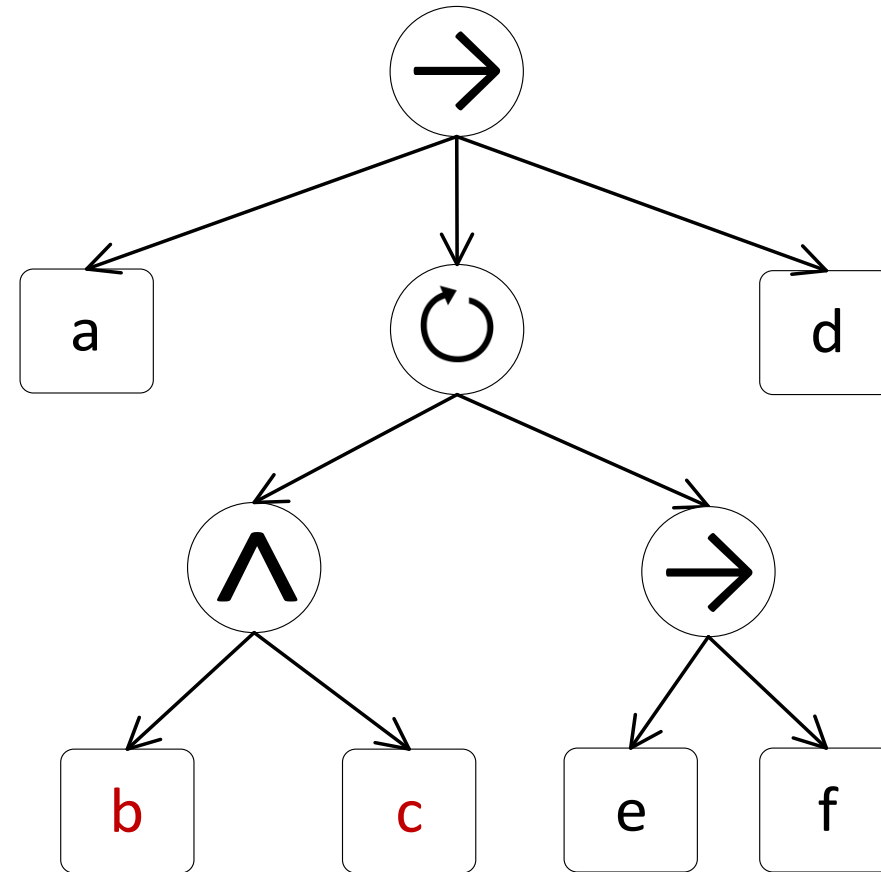
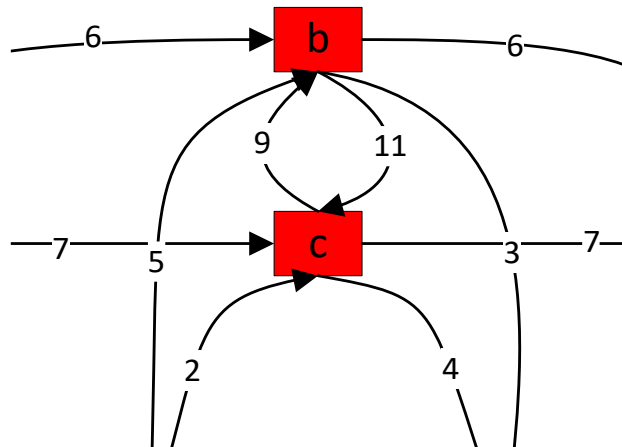
van der Aalst (2016), S. 222 ff.

Sequence Cut.



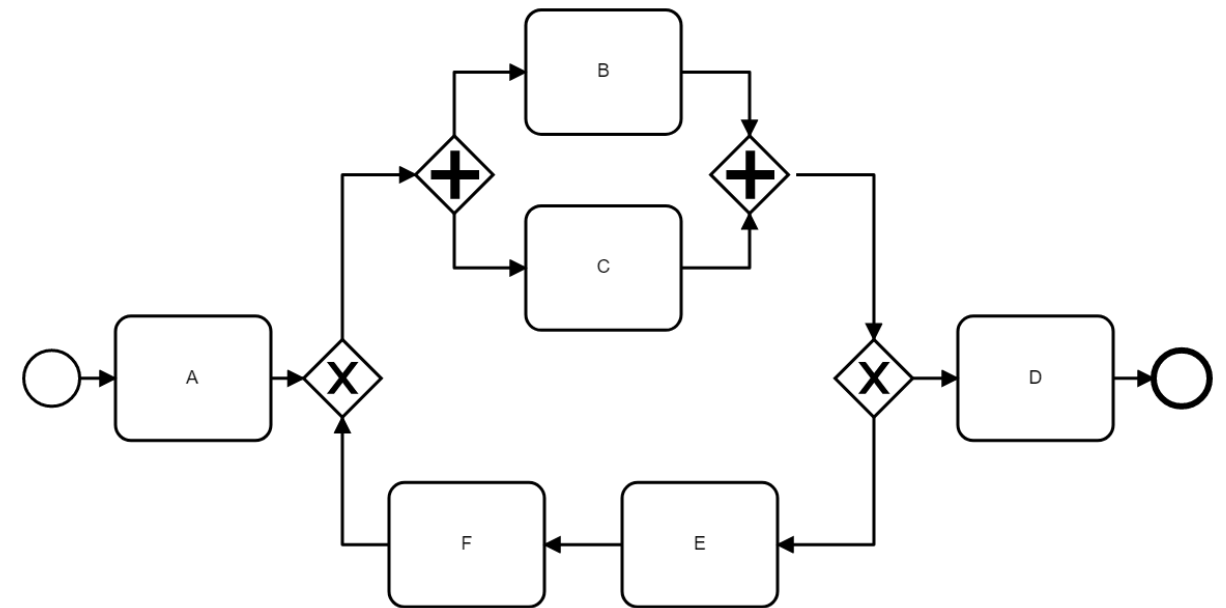
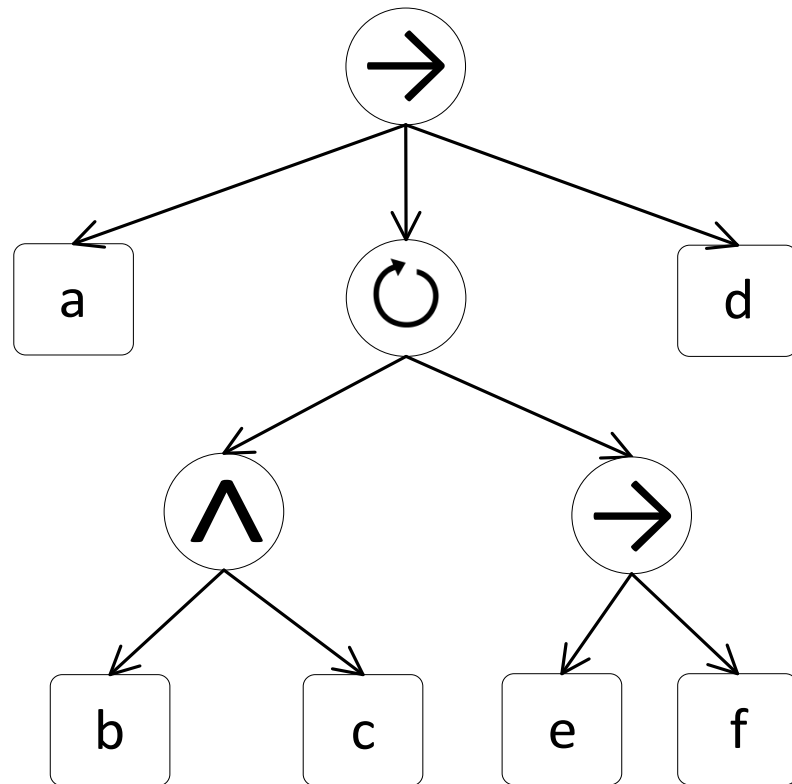
van der Aalst (2016), S. 222 ff.

Sequence Cut.



van der Aalst (2016), S. 222 ff.

Der Process Tree kann automatisch in ein wohlgeformtes BPMN-Modell transformiert werden.



van der Aalst (2016), S. 222 ff.

Hinweise zu den Übungsaufgaben.

- Bitte bilden Sie 2-3 Gruppen.
- Zu jeder der beiden Übungsaufgaben gibt es die Fallbeschreibung A und die Fallbeschreibung B.
- Nach der Bearbeitung der jeweiligen Übungsaufgabe stellen zwei Gruppen ihre Lösungsvorschlag dem Auditorium vor.
- Die Idee des Inductive Miners lernt man nur durch modellieren und diskutieren. Nutzen Sie die Möglichkeit zur Diskussion, auch wenn Ihnen Ihre Lösung nicht perfekt erscheint!
- Hinweis zur Bearbeitung:
 - Lesen Sie zunächst die komplette Aufgabe.
 - Skizzieren Sie Ihre Lösung auf einem A4-Blatt.
 - Übertragen Sie Ihre Lösung in eine PowerPoint-Folie.



Übungsaufgabe 2 (Anwenden des Inductive Miner)

Gruppe A

$$L_A = \begin{bmatrix} \langle a, c, d \rangle^{45}, \\ \langle b, c, d \rangle^{42}, \\ \langle a, c, e \rangle^{38}, \\ \langle b, c, e \rangle^{42} \end{bmatrix}$$

Gruppe B

$$L_B = \begin{bmatrix} \langle a, b, d \rangle^3, \\ \langle a, b, c, b, d \rangle^2, \\ \langle a, b, c, b, c, b, d \rangle^1 \\ \langle a, b, c, d \rangle^1 \end{bmatrix}$$

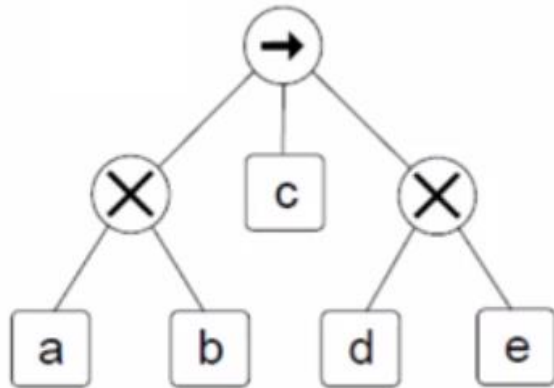
- Erzeugen Sie aus dem gegebenen Log einen Process Tree.
- Zusatz: Bilden Sie den Process Tree in einem BPMN-Prozessmodell ab. Diskutieren Sie mögliche Stärken und Schwächen Ihres Modells im Hinblick auf das zugrundeliegende Event-Log.



Übungsaufgabe 2 (Anwenden des Inductive Miner)

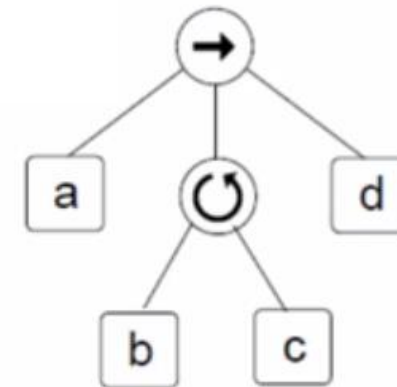
Gruppe A

$$L_A = \begin{bmatrix} \langle a, c, d \rangle^{45}, \\ \langle b, c, d \rangle^{42}, \\ \langle a, c, e \rangle^{38}, \\ \langle b, c, e \rangle^{42} \end{bmatrix}$$



Gruppe B

$$L_B = \begin{bmatrix} \langle a, b, d \rangle^3, \\ \langle a, b, c, b, d \rangle^2, \\ \langle a, b, c, b, c, b, d \rangle^1 \\ \langle a, b, c, d \rangle^1 \end{bmatrix}$$



Inhalte des Workshops

1

Wozu Process Mining und mit welcher Datenbasis?

2

Heuristic Miner

3

Inductive Miner

4

Weitere Themenfelder im Process Mining

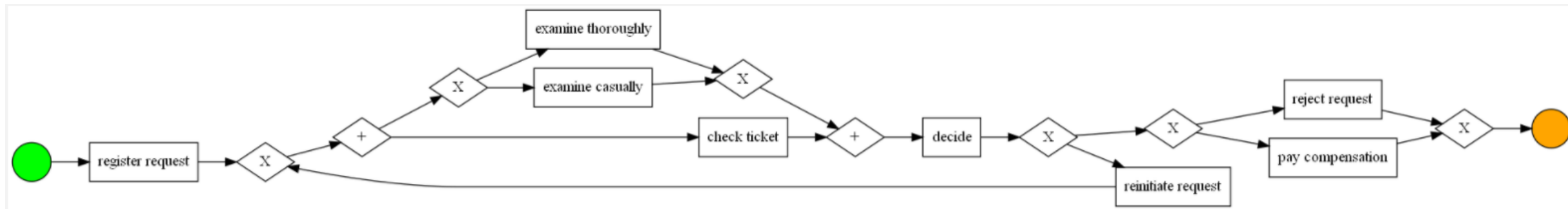
- **Fitness**
 - Das Modell kann das beobachtete Verhalten im Log vollständig wiedergeben.
- **Precision**
 - Das Modell kann kein anderes Verhalten, als das beobachtete Verhalten im Log wiedergeben.
- **Simplicity**
 - Das Modell ist leicht verständlich (endliche Menge an Knoten und Kanten, wohlgeformtes Prozessmodell)
- **Generalization**
 - Es handelt sich um ein generelles Modell, dass nicht zum Overfitting neigt.

- Event-Logs
 - „Object centric“ statt „process centric“
 - Unvollständig
 - Ausreißer
 - Logs in verschiedenen Informationssystemen haben unterschiedliche Detaillierungsgrade
- Cross Organizational Process Mining
- Concept Drift
- Erzeugen qualitativ hochwertige Prozessmodelle
- ...

van der Aalst (2016), S. 449 f.; van der Aalst et al. (2012)

- Celonis
- Disco
- ProM
- PM4Py
- ...

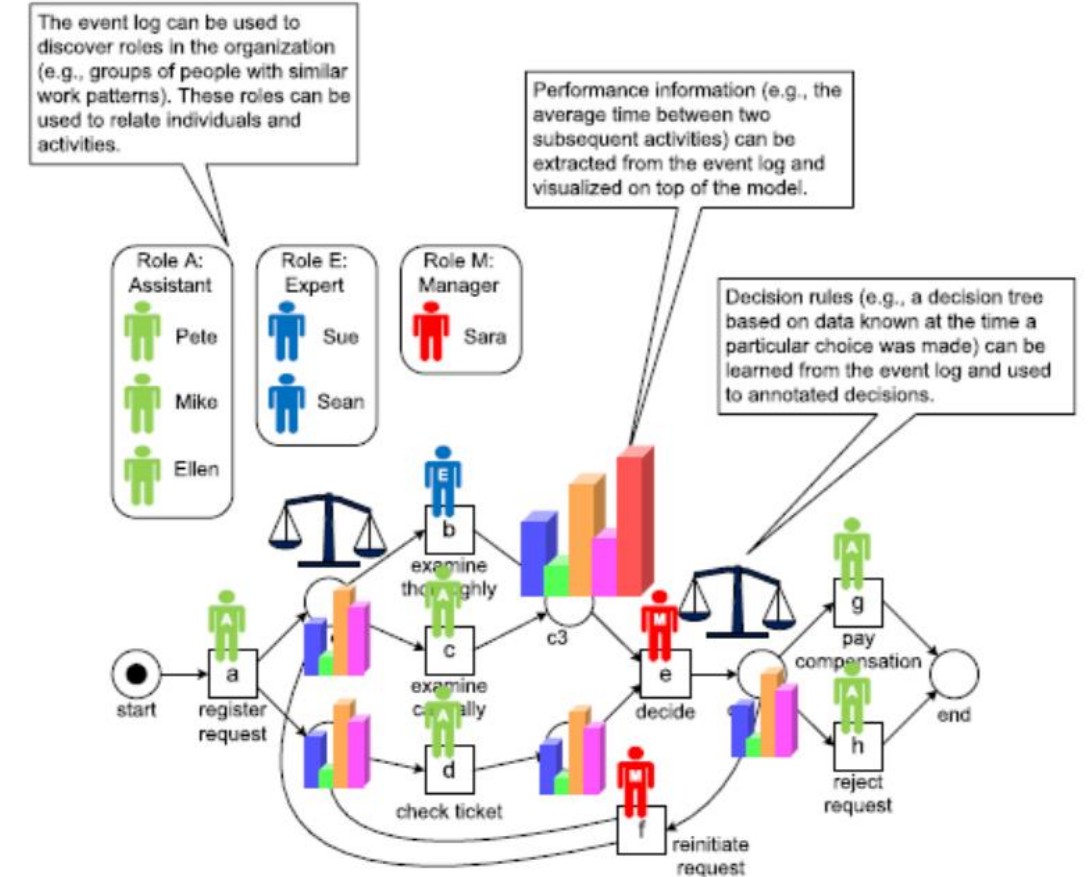
index ▲	Activity	Costs	Resour...	case:c...	case:creator	concept:name	org:re...	time:timestamp
0	register request	50	Pete	3	Fluxicon Nitro	register request	Pete	2010-12-30T13:32:00.000
1	examine casually	400	Mike	3	Fluxicon Nitro	examine casually	Mike	2010-12-30T14:06:00.000
2	check ticket	100	Ellen	3	Fluxicon Nitro	check ticket	Ellen	2010-12-30T15:34:00.000
3	decide	200	Sara	3	Fluxicon Nitro	decide	Sara	2011-01-06T08:18:00.000
4	reinitiate request	200	Sara	3	Fluxicon Nitro	reinitiate request	Sara	2011-01-06T11:18:00.000
5	examine thoroughly	400	Sean	3	Fluxicon Nitro	examine thoroughly	Sean	2011-01-06T12:06:00.000
6	check ticket	100	Pete	3	Fluxicon Nitro	check ticket	Pete	2011-01-08T10:43:00.000
7	decide	200	Sara	3	Fluxicon Nitro	decide	Sara	2011-01-09T08:55:00.000
8	pay compensation	200	Ellen	3	Fluxicon Nitro	pay compensation	Ellen	2011-01-15T09:45:00.000
9	register request	50	Mike	2	Fluxicon Nitro	register request	Mike	2010-12-30T10:32:00.000
10	check ticket	100	Mike	2	Fluxicon Nitro	check ticket	Mike	2010-12-30T11:12:00.000
11	examine casually	400	Sean	2	Fluxicon Nitro	examine casually	Sean	2010-12-30T13:16:00.000
12	decide	200	Sara	2	Fluxicon Nitro	decide	Sara	2011-01-05T10:22:00.000



Viner et al. (2020)
<https://www.processmining-software.com/tools/>

Weitere Anwendungsfelder im Process Mining

- Conformance Checking
- Bottleneck Analysis und Basis von Prozessverbesserungen
- Decision Mining
- Mining von weiteren Prozessperspektiven (Datenmodelle, Ressourcen, Rollen, ...)
- ...



Di Francescomarino et al. (2018); van der Aalst (2016), S. 40 f.

Process Mining

- Die Idee von Process Mining ist das “Entdecken”, Überwachen und Verbessern von realen (und nicht angenommenen) Prozessen durch die Extraktion von Wissen aus Event-Logs.

Event-Log

- Basis für die Process Discovery.
- Case-ID, Aktivitätsname und Zeitstempel sind notwendige Informationen.
- Weitere Attribute, wie Ressourcen, sind möglich.
- Verschiedene IT-Systeme (ERP-System, WfMS, Bestellsystem) können Event-Logs erzeugen.

Verfahren zur Rekonstruktion des Kontrollflusses

- Heuristic Miner erzeugt ein Casual Net.
- Der Detaillierungsgrad des Casual Net kann durch Parameter verändert werden.
- Inductive-Miner erzeugt Process Trees.
- Process Trees lassen sich in wohlgeformte Prozessmodelle transformieren.

Tobias Seyffarth

✉ tobias@seyffarth.me

🔗 <https://seyffarth.me>

🐙 <https://github.com/tobiasseyffarth>

R^G https://www.researchgate.net/profile/Tobias_Seyffarth

- Di Francescomarino, Chiara; Ghidini, Chiara; Maggi, Fabrizio Maria; Milani, Fredrik (2018): Predictive Process Monitoring Methods: Which One Suits Me Best? In: *International Conference on Business Process Management (BPM)*, S. 462–479.
- Funk, Burkhardt (2010): Geschäftsprozessintegration mit SAP. Fallstudien zur Steuerung von Wertschöpfungsprozessen entlang der Supply Chain. Berlin, Heidelberg.
- van der Aalst, Wil; Adriansyah, Arya; Medeiros, Ana Karla Alves de; Arcieri, Franco; Baier, Thomas; Blickle, Tobias et al. (2012): Process Mining Manifesto. In: *Business Process Management Workshops 99*, S. 169–194. DOI: 10.1007/978-3-642-28108-2_19.
- van der Aalst, Wil (2016): Process Mining. Data Science in Action. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Viner, Daniel; Stierle, Matthias; Matzner, Martin (2020): A Process Mining Software Comparison. In: *2nd International Conference on Process Mining (ICPM)*.
- van der Aalst, Wil: Business Process Intelligence (BPI).
https://www.youtube.com/playlist?list=PLG_1ZxIPXO0uA-LolJSQH2jzJ8oUiq9hu
- Dumas, Marlon: Business Process Mining Course.
<https://www.youtube.com/watch?v=LrEYPx0Rqyk&list=PL2LDY1TRcen3jy3W3D1gl5yKxnd7tMEBT>