Технологии Data Mining & Process Mining

MEDIA-TEL by devoteam

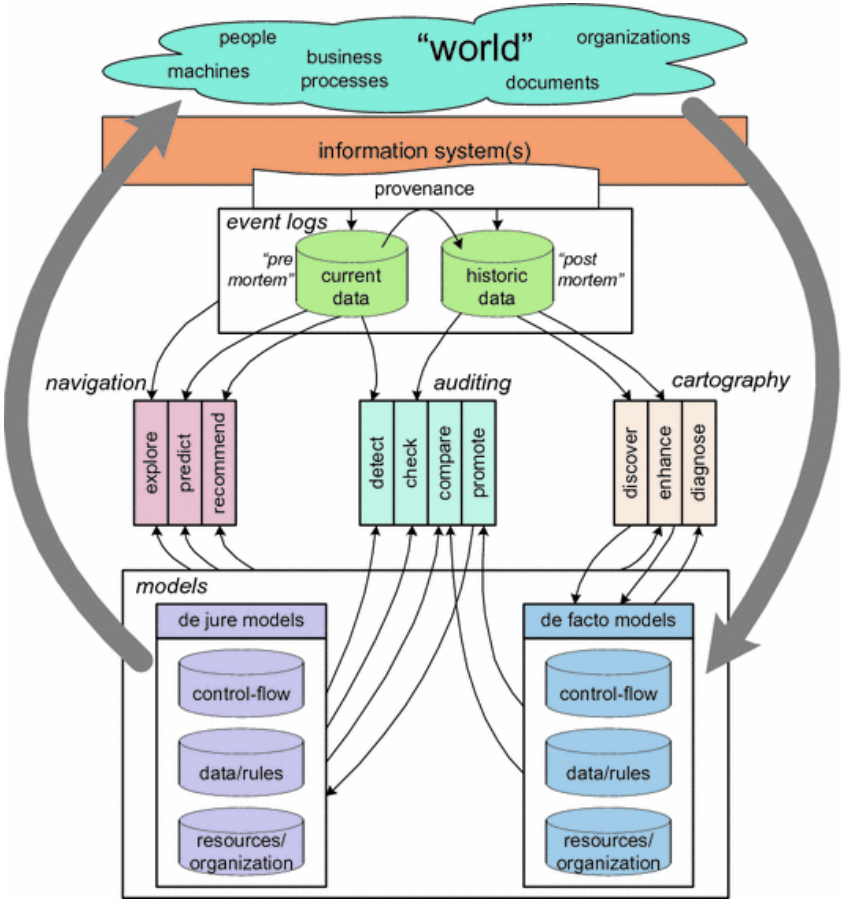
24 March 2021

План

1. Введение в Process Mining
    1. Типы задач которые можно решить при помощи Process Mining vs Data Mining
    2. Данные, необходимые для Process Mining
    3. Принципы получения моделей процессов из лога событий
    4. Нотации для описания и визуализации процессов (DFG, BPMN, Petri nets и др.)
    5. Работа с моделью процесса:
        1. Генерация вариантов исполнения процесса согласно заданной модели(play-out)/симуляция,
        2. Получение модели процесса по имеющимся событиям (play-in/process discovery),
        3. Наложение записей о событиях на модель, выявление несоответствий реальных данных модели, анализ производительности (replay/log alignment/conformance checking/performance analysis).

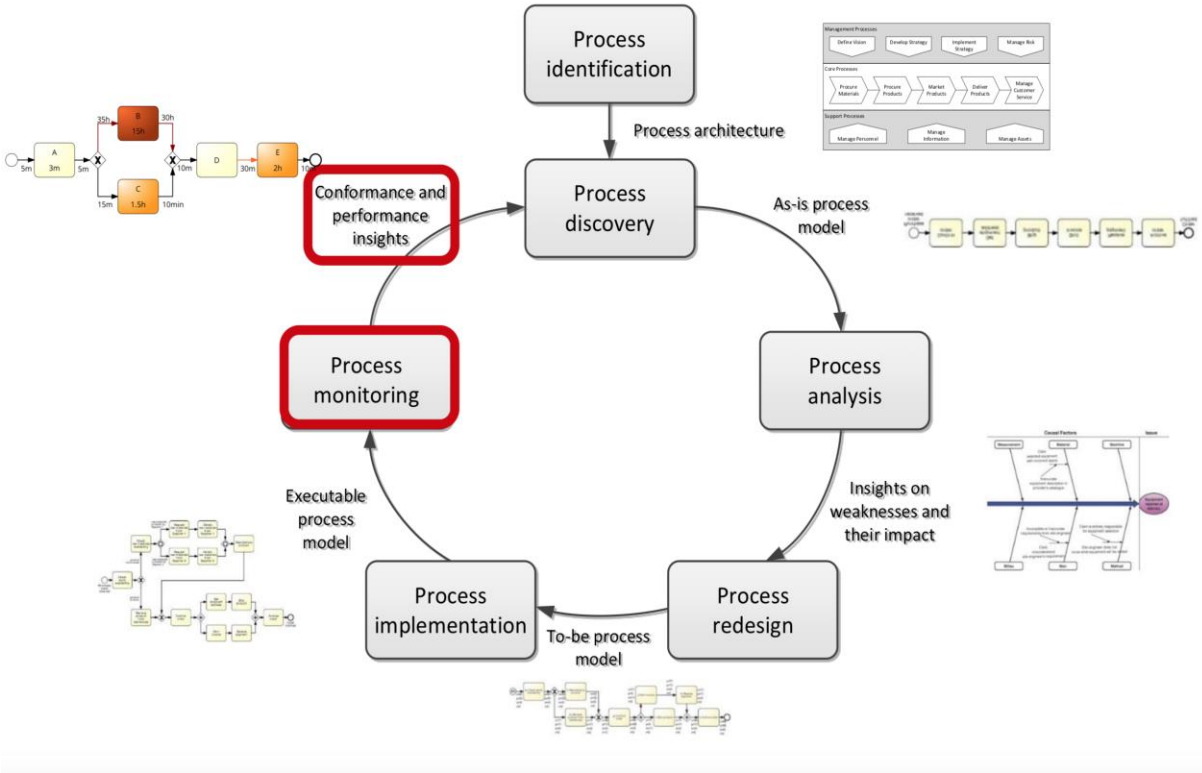
MEDIA-TEL by devoteam

“Экосистема” Process Mining



MEDIA-TEL
by devoteam

Этапы применения Process Mining



MEDIA-TEL by devoteam

Вопросы на которые может помочь ответить Process Mining

- Какому процессу на самом деле следуют люди, по сравнению с предполагаемым процессом?
- Где узкие места в выполняющихся процессах?
- В каких точках процесса люди или другие участники взаимодействия отклоняются от предопределённого процесса?
- Как выглядят самые частые пути прохождения процесса - главный сценарий выполнения (happy path)?
- Какие факторы влияют на возникновение узкого места при выполнении процесса?
- Можно ли предсказать проблемы (задержку, отклонение, риски) для процессов до их завершения?
- Можно ли предложить меры по предотвращению проблем?
- Как лучше преобразовать процессы в организации?
- …

Данные, необходимые для Process Mining

- Получение модели процесса возможно из любых данных, упорядоченных во времени, образующихся в результате работы и взаимодействия людей, машин, программного обеспечения, если из данных можно получить следующие переменные:
- дата/время события (**timestamp**) - обязательный
- действие/активность (**activity_id**) - обязательный
- идентификатор конкретной цепочки событий в рамках одного экземпляра процесса (**case_id**) - обязательный
- индикатор жизненного цикла (**lifecycle_id**) действия/активности - условно опциональный (может быть сформирован искусственно для простых случаев)
- идентификатор экземпляра активности (**activity_instance_id**) - условно опциональный
- ресурс (**resource_id**), используемый при выполнении действия/активности - опциональный
- прочие атрибуты (стоимость, производительность и т.п.) - опционально

MEDIA-TEL by devoteam

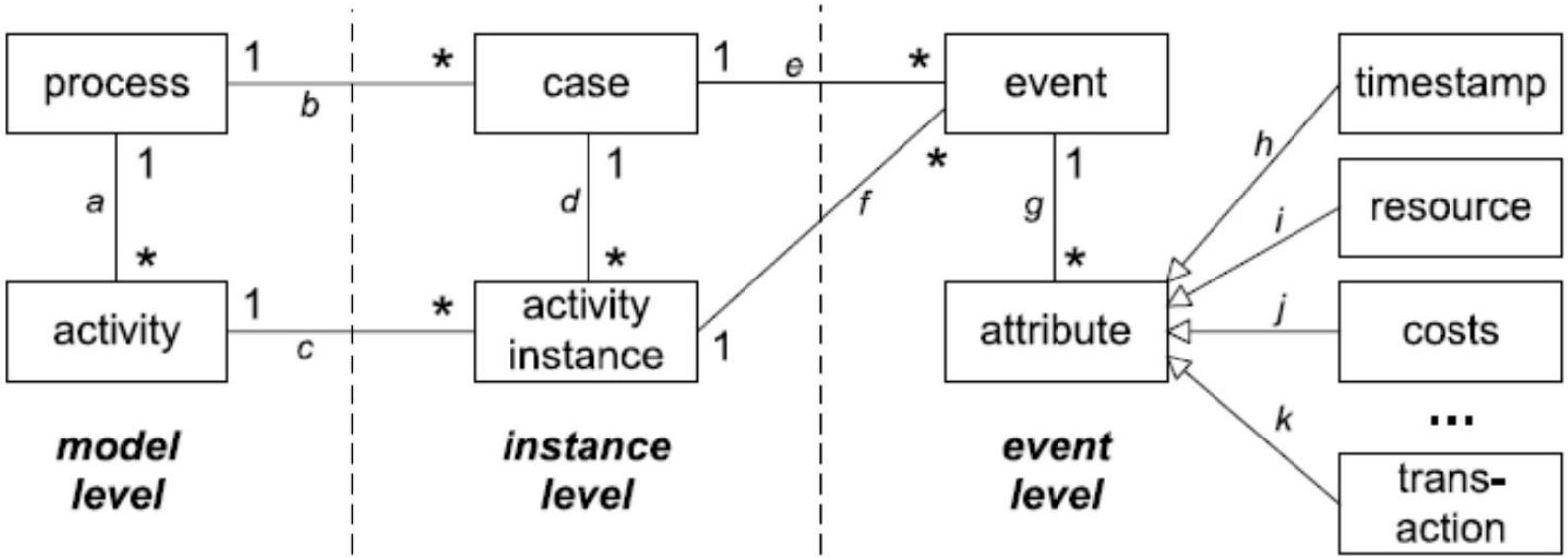
Данные, необходимые для Process Mining

	case_id	activity_id	resource_id	lifecycle_id	activity_instance_id	timestamp
1:	1	A	R1	complete	1	2020-02-17 11:00:00
2:	1	B	R1	complete	2	2020-02-17 11:10:00
3:	2	B	R1	complete	3	2020-02-17 10:50:00
4:	2	A	R1	complete	4	2020-02-17 11:05:00
5:	1	C	R1	complete	5	2020-02-17 11:20:00
6:	2	C	R1	complete	6	2020-02-17 11:35:00

	case_id	activity_id	resource_id	lifecycle_id	activity_instance_id	timestamp
1:	1	A	R1	start	1	2020-02-17 10:00:00
2:	1	A	R1	complete	1	2020-02-17 11:00:00
3:	1	B	R1	start	2	2020-02-17 10:30:00
4:	1	B	R1	complete	2	2020-02-17 11:10:00
5:	2	B	R1	start	3	2020-02-17 10:10:00
6:	2	B	R1	complete	3	2020-02-17 10:50:00
7:	2	A	R1	start	4	2020-02-17 10:05:00
8:	2	A	R1	complete	4	2020-02-17 11:05:00
9:	1	C	R1	start	5	2020-02-17 11:05:00
10:	1	C	R1	complete	5	2020-02-17 11:20:00
11:	2	C	R1	start	6	2020-02-17 11:15:00
12:	2	C	R1	complete	6	2020-02-17 11:35:00

MEDIA-TEL by devoteam

Модель данных



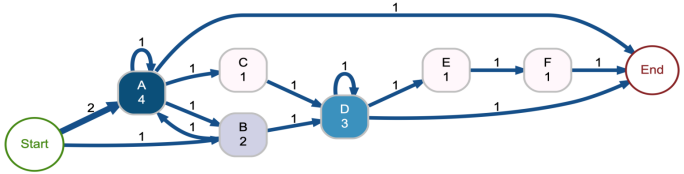
MEDIA-TEL
by devoteam

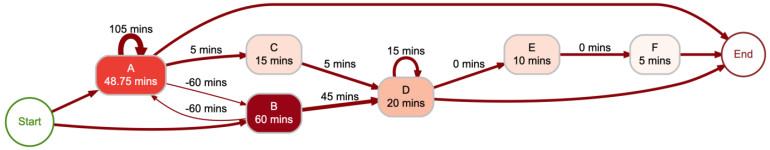
Принципы получения моделей процессов из лога событий - Process Discovery

	case_id	activity_id	resource_id	lifecycle_id	activity_instance_id	timestamp
1:	1	A	R1	complete	1	2020-02-17 11:00:00
2:	1	B	R1	complete	2	2020-02-17 11:00:00
3:	2	B	R1	complete	3	2020-02-17 11:00:00
4:	2	A	R1	complete	4	2020-02-17 11:00:00
5:	1	D	R1	complete	5	2020-02-17 12:15:00
6:	1	D	R1	complete	6	2020-02-17 12:45:00
7:	2	A	R2	complete	7	2020-02-17 13:15:00
8:	3	A	R2	complete	8	2020-02-17 12:00:00
9:	3	C	R2	complete	9	2020-02-17 12:20:00
10:	3	D	R2	complete	10	2020-02-17 12:40:00
11:	3	E	R2	complete	11	2020-02-17 12:50:00
12:	3	F	R1	complete	12	2020-02-17 12:55:00

Traces:
1: ABDD
2: BAA
3: ACDEF

- Alpha miner
- Inductive miner
- Heuristics miner
- Fuzzy miner
- Directly-follows miner
- Social Network miner
- Evolutionary Tree Miner
- Split Miner
- …





MEDIA-TEL by devoteam

Принципы получения моделей процессов из лога событий - Process Discovery

Direct Succession: x>y, если, в каких-то случаях, за активностью x следует y.
Causality: x->y, если x>y, но не y>x.
Parallel: x||y, если x>y и y>x.
Unrelated: x#y, если !(x>y) и !(y>x)

Traces:
1: ABDD
2: BAA
3: ACDEF
4: ABDD
5: ACDEF
6: ABDD

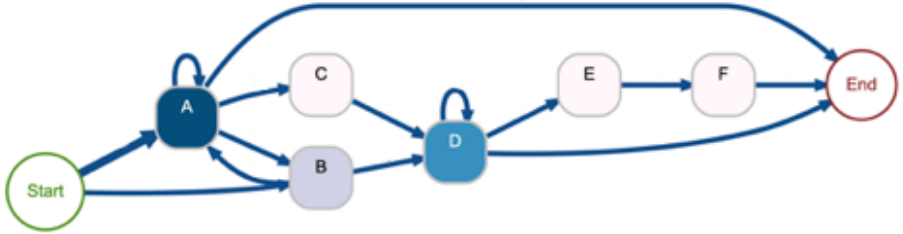
L1 = [<A,B,D,D>^3, <A,C,D,E,F>^2, <B,A,A>]

A>B
B>D
D>D
B>A
A>A
A>C
C>D
D>E
E>F

A||B
B||A

B->D
D->D
A->A
A->C
C->D
D->E
E->F

	A	B	C	D	E	F
A	<->	||	->	#	#	#
B	||	#	#	->	#	#
C	<-	#	#	->	#	#
D	#	<-	<-	<->	-> —>>	#
E	#	#	#	<-	#	->
F	#	#	#	#	<-	#



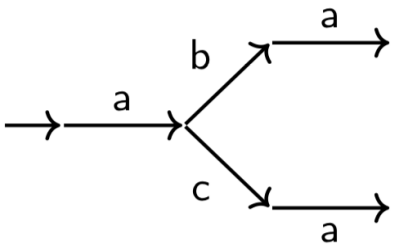
MEDIA-TEL
by devoteam

Нотации для описания и визуализации процессов

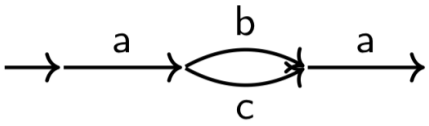
- Deterministic/non-Deterministic Finite State Automata
- Petri Nets
- YAWL (Yet Another Workflow Language)
- BPMN (Business Process Model and Notation)
- C-nets
- Process Trees
- …

MEDIA-TEL
by devoteam

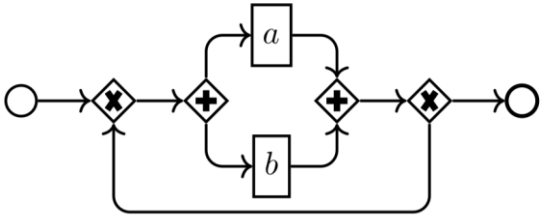
Нотации для описания и визуализации процессов

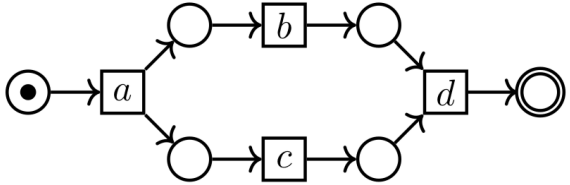


(a) DFA.



(b) Minimal DFA.





(a) start **(b)** end **(c)** concurrency **(d)** exclusive choice **(e)** inclusive choice

MEDIA-TEL by devoteam

Нотации для описания и визуализации процессов - Petri Nets

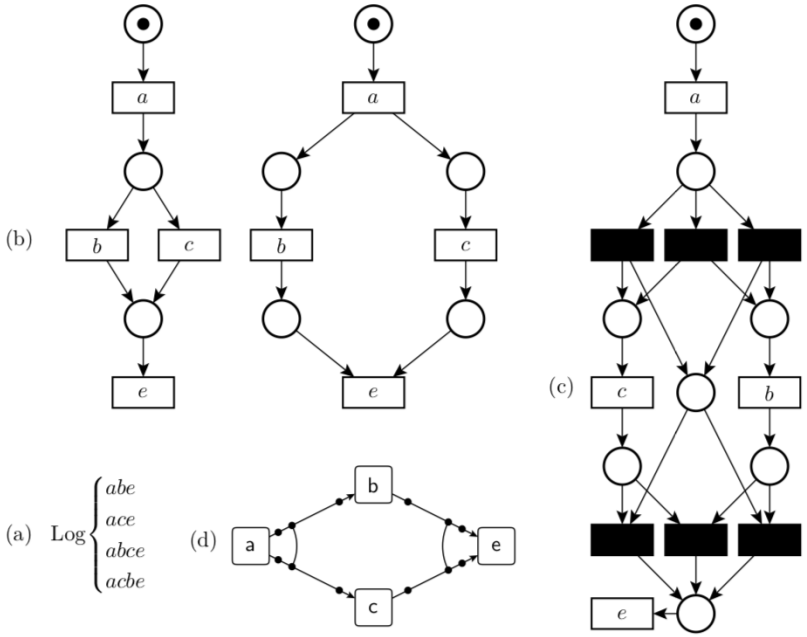


Fig. 1. (a) A log. (b) Set of two Petri nets describing the log. (c) Single Petri net describing the log with the help of silent transitions. (d) Causal net of the same log.

MEDIA-TEL by devoteam

Нотации для описания и визуализации процессов - Petri Net, BPMN, DFG

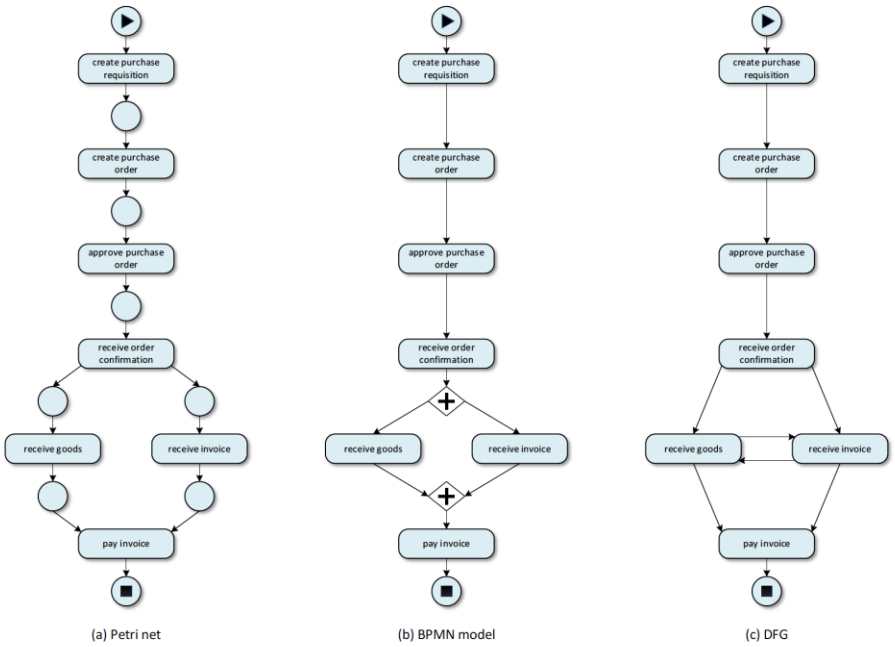
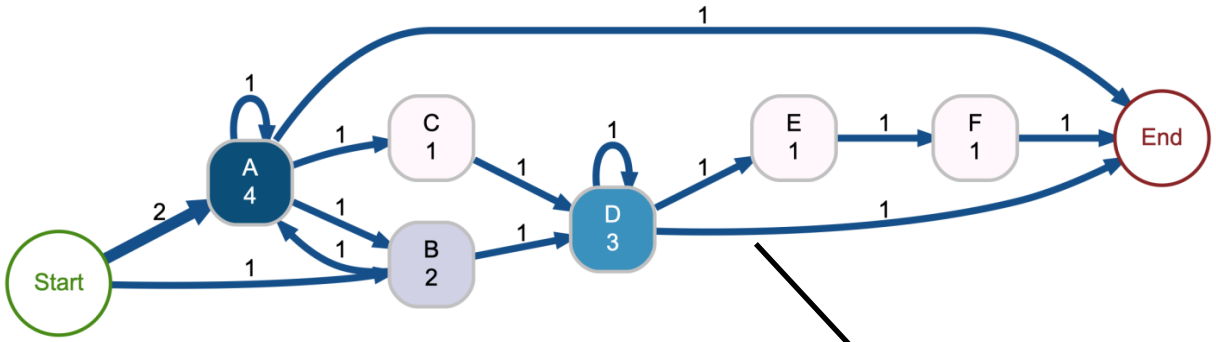


Fig. 1. Three process models discovered for the Purchase-to-Pay (P2P) process considering only the frequent "happy paths". The Petri net model (a) and the Business Process Model and Notation (BPMN) model (b) specify the same behavior. The Directly-Follows Graph (DFG) (c) allows for traces not allowed in the Petri net and BPMN model (e.g., a loop involving the activities *receive goods* and *receive invoice*).

Wil Van der Aalst, RWTH Aachen University: A Practitioner's Guide to Process Mining: Limitations of the Directly-Follows Graph

MEDIA-TEL by devoteam

Работа с моделью: Play-out



Traces:
ABDD
BAA
ACDEF
ABDEF
ABD
BD
BDD
BDEF

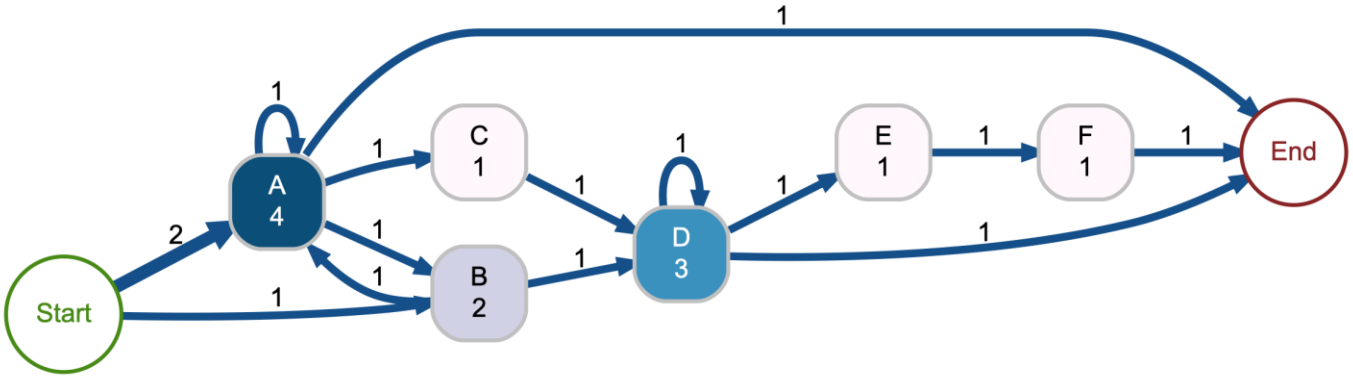
ACDD
A
AA
AACD
AACDED
AACDD
AACDDEF
AABABDED

ABACD
ABACDD
ABACDDE
D
ABACDEF
…

MEDIA-TEL
by devoteam

Работа с моделью: Play-in

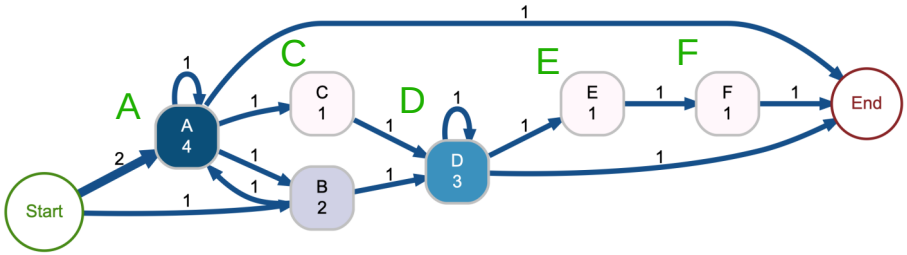
Traces:
ABDD
BAA
ACDEF
ABDEF
ABD
BD



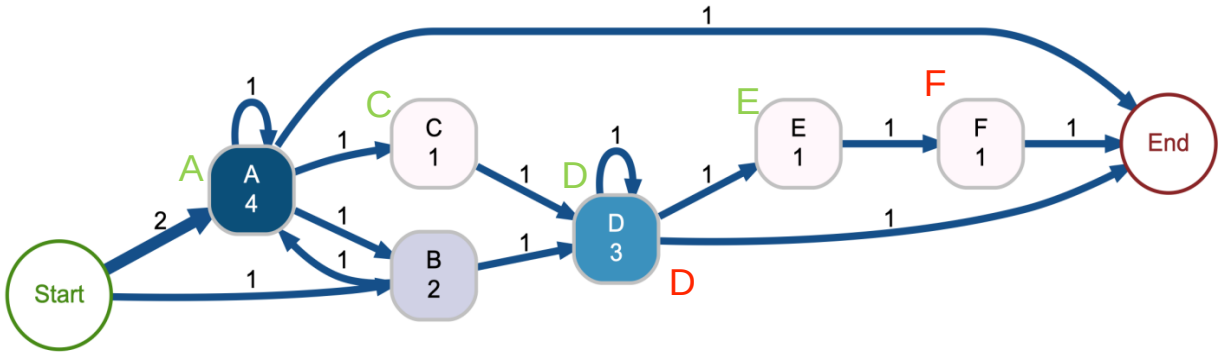
MEDIA-TEL
by devoteam

Работа с моделью: Replay/conformance checking

Traces:
1: ABDD
2: BAA
3: ACDEF



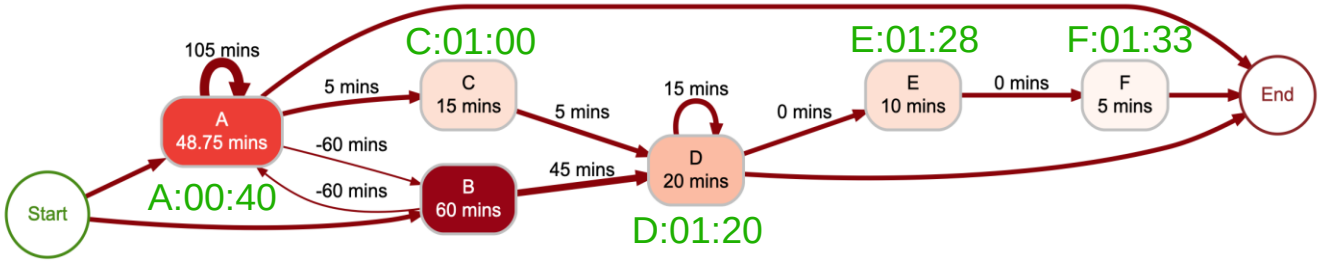
Trace: ACDEDF



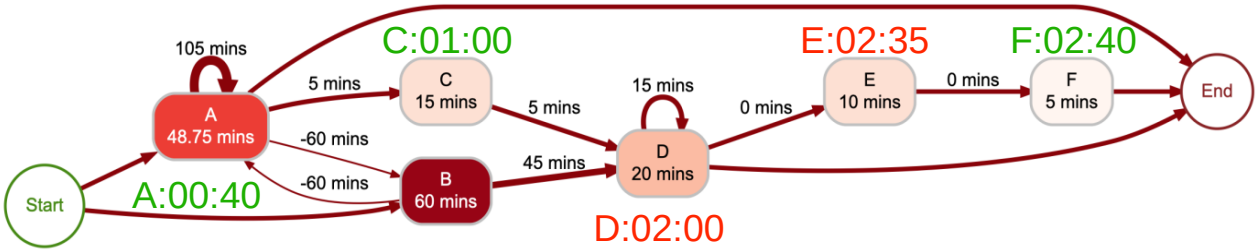
MEDIA-TEL
by devoteam

Работа с моделью: Replay/conformance checking + performance checking

Traces:
1: ABDD
2: BAA
3: ACDEF



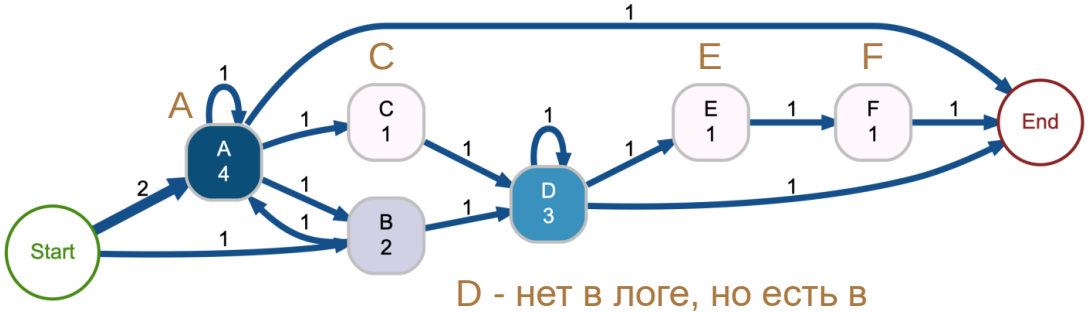
Traces:
1: ABDD
2: BAA
3: ACDEF



MEDIA-TEL by devoteam

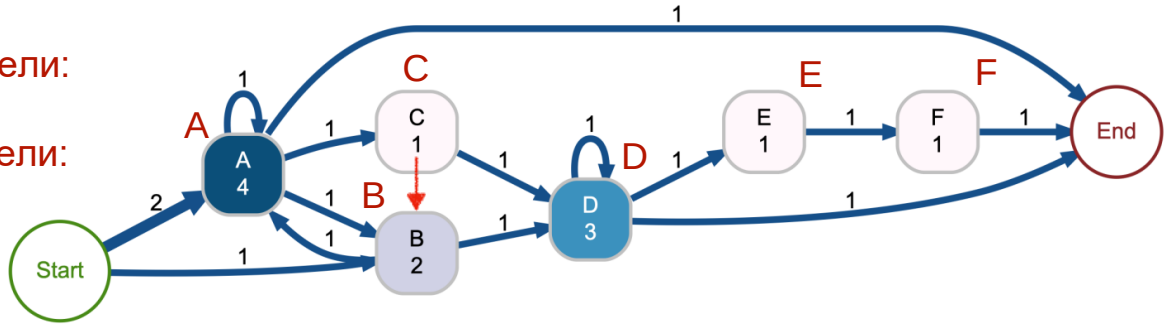
Работа с моделью: Conformance checking - alignments

Traces:
ABDD
BAA
ACDEF
ABDEF
ABD
BD
ACEF
ACBDEF



D - нет в логе, но есть в модели

1) в модели нет ребра CB
2) C есть в логе, но нет в модели: ABDEF
3) B есть в логе, но нет в модели: ACDEF



MEDIA-TEL by devoteam

Качество модели: метрики и Conformance Checking

- Алгоритмы, используемые для получения модели из логов событий имеют недостатки: модель может описывать поведение, которое на самом деле отсутствует, а также может исключать поведение, которое на самом деле имеет место.
- В связи с этим несовершенством возникает потребность в оценке качества полученной модели.
- Для оценки качества есть несколько методик оценки и метрик.
- Методика log-conformance checking позволяет сравнить расхождение между логом и полученной моделью.
- Методика model-conformance checking позволяет сравнить полученную модель с эталонной моделью, если такая имеется. Также можно сравнивать модели, полученные на разных логах событий, полученных из той же системы.

Process Discovery: Критерии качества модели

- Качественные характеристики:

  - Soundness (добротность) - модель должна удовлетворять нескольким критериям (о них чуть позже).

  - Fitness/recall (годность) - характеристика, отражающая степень, с которой полученная модель соответсвует поведению, наблюдаемому в логе событий.
  - Precision (точность) - характеристика, отражающая степень, с которой лог событий, генерируемый согласно модели процесса не допускает поведение, которое не встречается в реальном логе событий.
  - Generalization (обобщение) - оценивает степень, с которой модель сможет воспроизвести будущее, ещё не наблюдавшееся поведение.
  - Simplicity (простота) - насколько простая модель.
  - Rediscoverability - характеристика, отражающая степень, с которой полученная из логов модель соответствует исходной модели системы (критерий качества алгоритма получения модели).

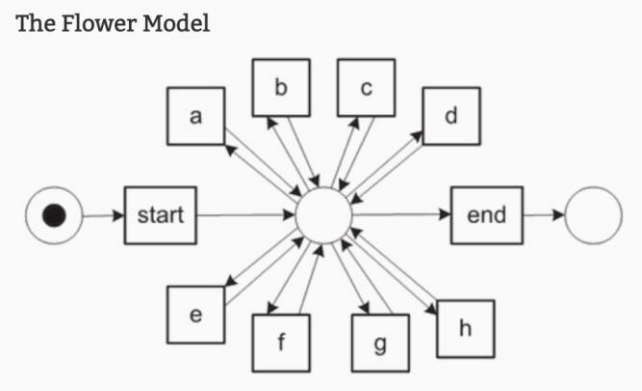
MEDIA-TEL
by devoteam

Критерии качества модели - добротность (soundness)

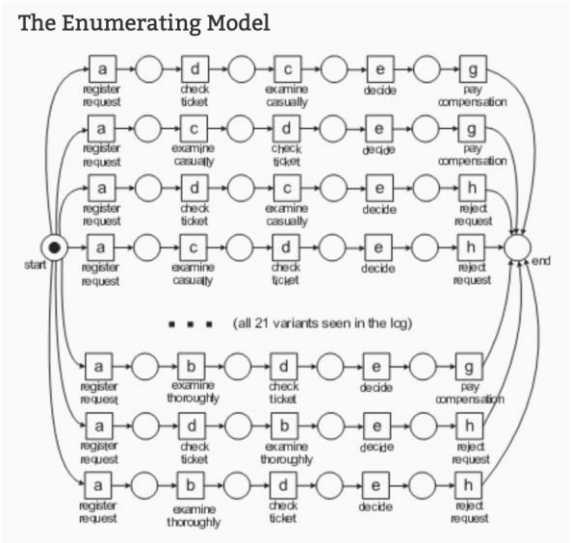
Модель является добротной (sound), только если выполняются следующие четыре свойства (в терминах Petri net):

1. “Места должны быть безопасными”. Максимум один “токен” должен быть в каком-либо “месте” в любой момент времени.
2. Должно быть корректное завершение, то есть, как только токен попадает в конечное “место”, все остальные “места” должны быть пустыми.
3. Должна быть возможность завершения. Что бы мы ни делали, мы всегда должны иметь возможность достигнуть состояния с единственным “токеном” в завершающем “месте”.
4. Не должно быть “мёртвых” частей. Все части модели процесса должны быть потенциально активируемые. Должны быть сценарии, при которых все части играют роль.

Критерии качества модели - примеры



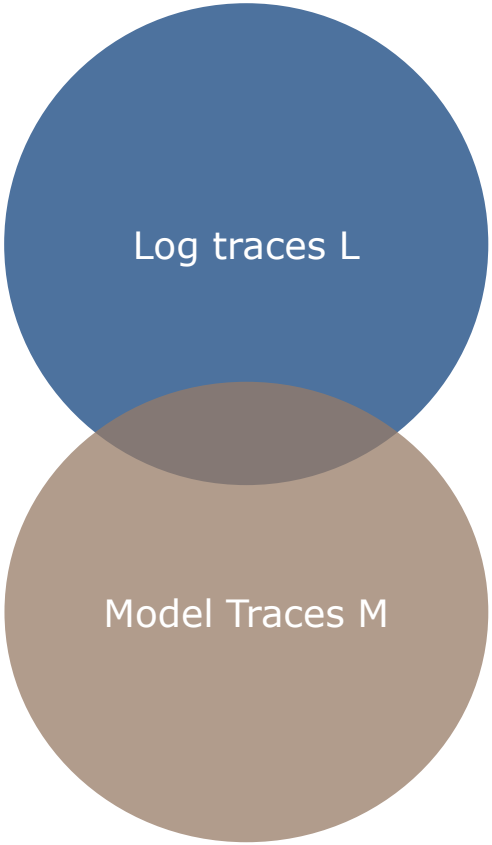
- очень простая модель
- отличный fitness - любой возможный trace можно воспроизвести
- отличный generalization - модель позволяет любое поведение
- очень низкий precision - модель позволяет многое, чего не было в исходных логах



- модель отражает только то что было в логах
- хорошая fitness и precision
- полностью отсутсвует generalization - не допускается поведение, которого не было в логах
- модель достаточно сложная (simplicity неудовлетворительная)

MEDIA-TEL by devoteam

Критерии качества модели - Precision & Fitness



Precision:

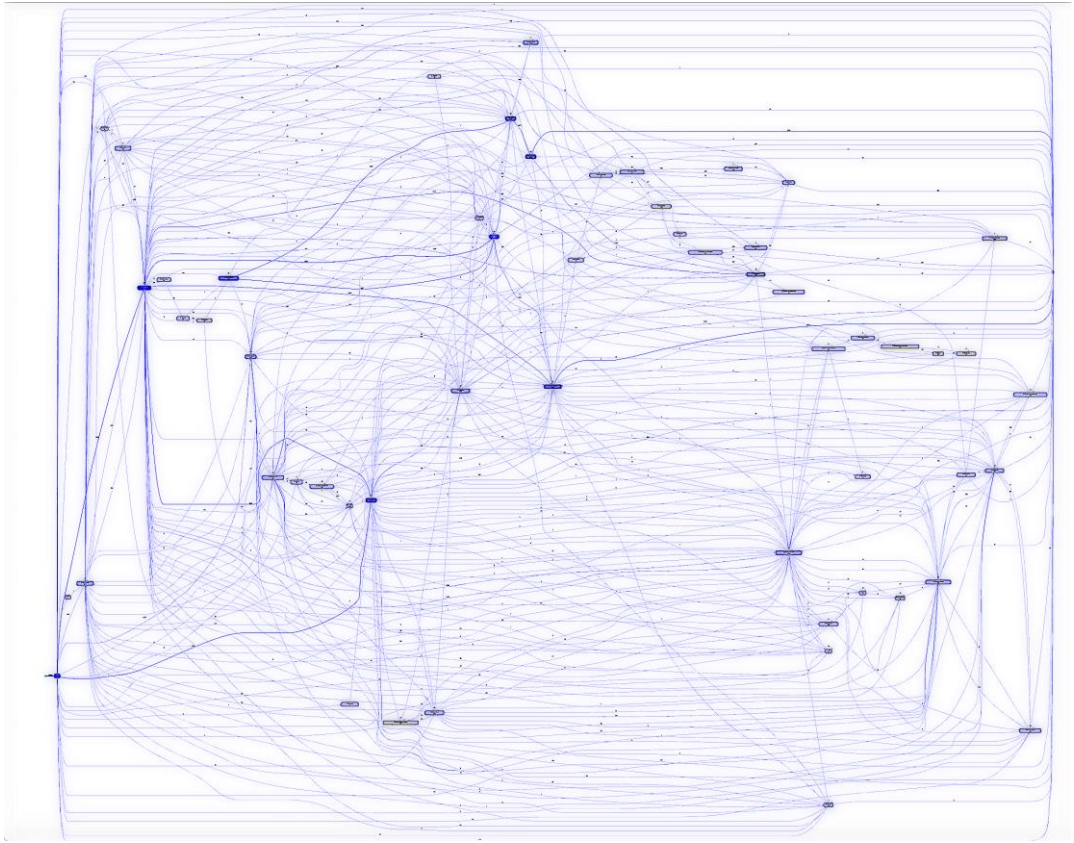
$$\frac{|L \cap M|}{|M|}$$

Recall (Fitness):

$$\frac{|L \cap M|}{|L|}$$

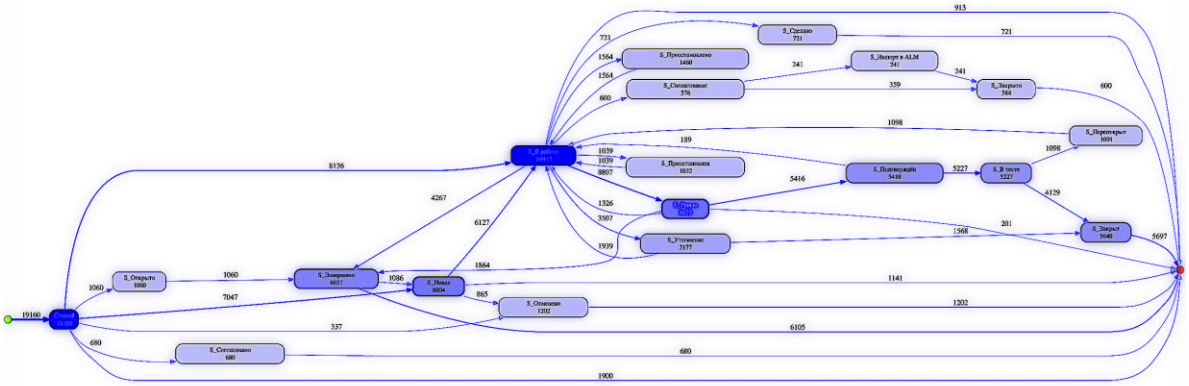
MEDIA-TEL
by devoteam

Сложность модели и варианты упрощения: процесс типа “спагетти”.



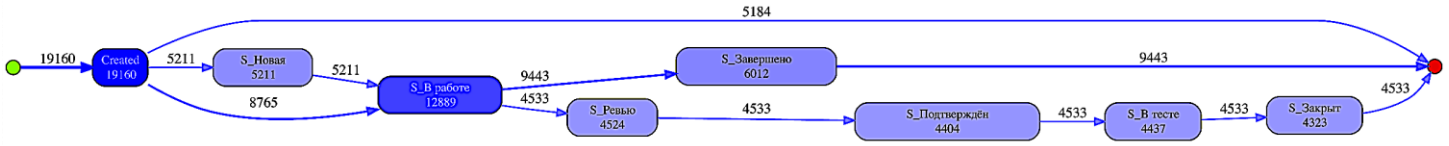
MEDIA-TEL
by devoteam

Сложность модели и варианты упрощения: процесс типа “лазанья”.



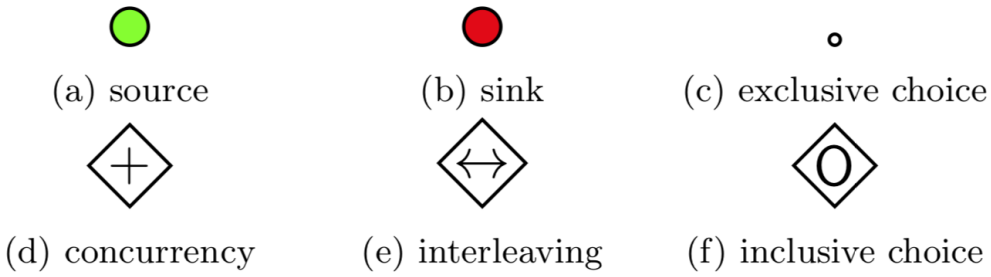
MEDIA-TEL by devoteam

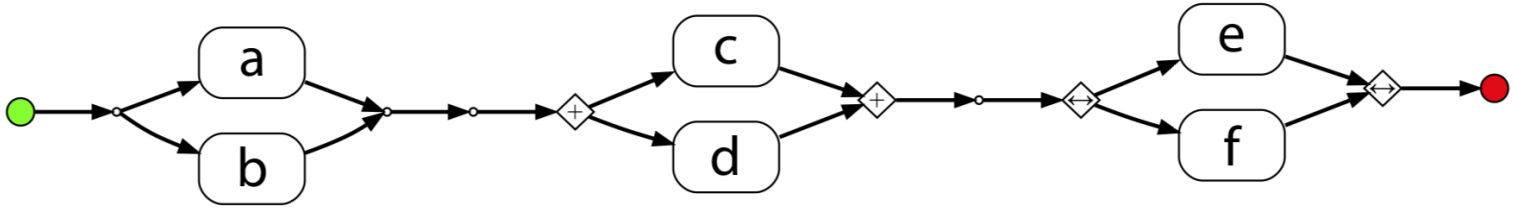
Сложность модели и варианты упрощения: процесс типа “три макаронины”.



MEDIA-TEL by devoteam

Inductive Visual Miner constructs





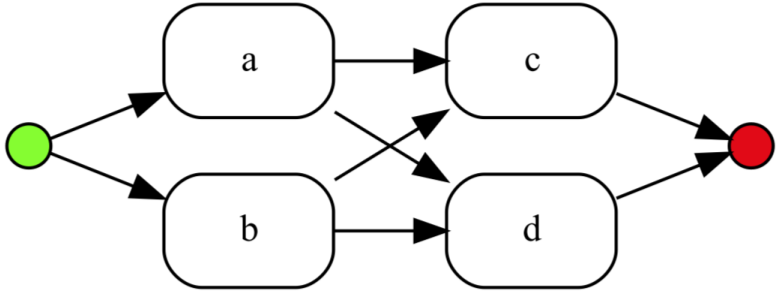
MEDIA-TEL by devoteam

Directly-Follows Miner constructs

(a) source

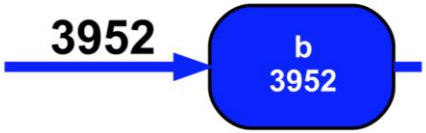
(b) sink

(c) exclusive choice

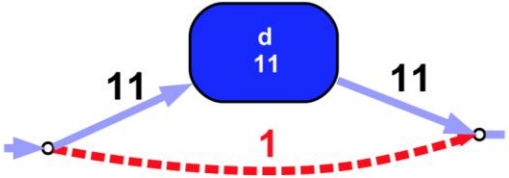


MEDIA-TEL
by devoteam

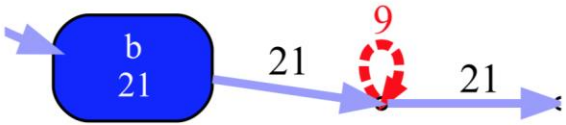
Визуализация отклонений



(a) Edge and activity.



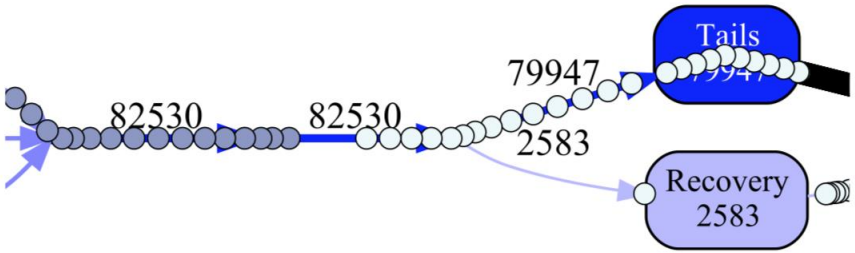
(b) Model move.

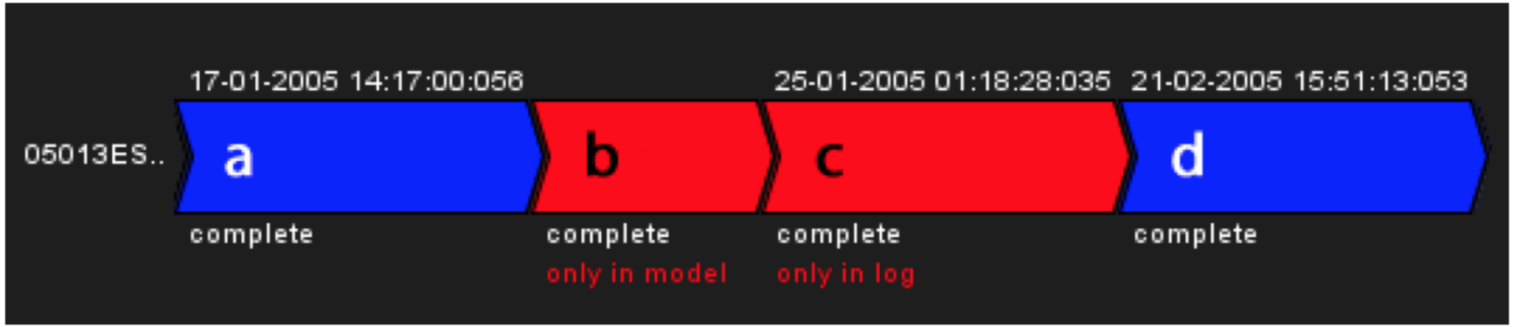


(c) Log move.

MEDIA-TEL
by devoteam

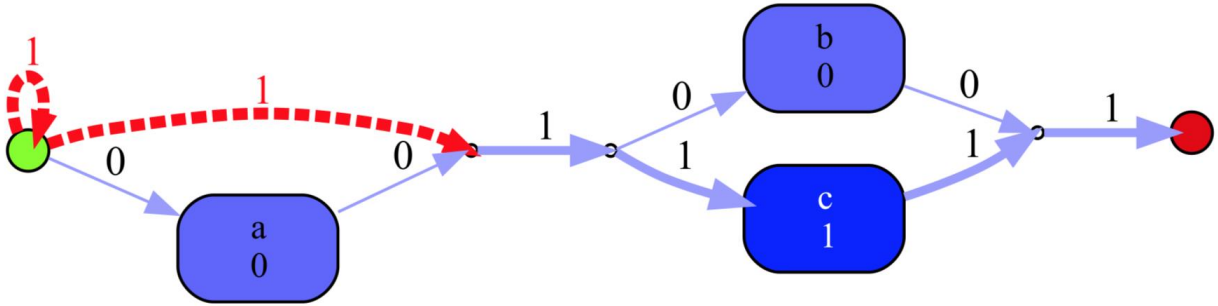
Анимация, трассировка





MEDIA-TEL
by devoteam

Интерпретация отклонений



Trace: t = ⟨b, c⟩

$$M = \text{sequence}$$
$$a \quad \text{xor}$$
$$b \quad c$$

trace	b	-	c
model	-	a	c

- Согласно модели, событие *a* должно было быть выполнено, однако в трейсе такого события не было найдено. Красная дуга отражает это отклонение.
- Согласно трейсу, событие *b* должно было произойти, но модель не поддерживает такого события в данном месте. Это событие считается излишним с точки зрения модели и отображается петлёй.

MEDIA-TEL by devoteam

Неоднозначность отклонений

Не всегда можно однозначно интерпретировать отклонения, несмотря на интуитивно понятную визуализацию. Например, следующее “выравнивание” трейса и модели, имеет такое же количество отклонений:

trace	-	b	c
model	a	b	-

В этом “выравнивании”, согласно трейсу, событие *c* должно было произойти вместо *b.*

Таким образом, несколько разных “выравниваний” могут считаться оптимальными и этот факт нужно учитывать при интерпретации отклонений.

MEDIA-TEL by devoteam

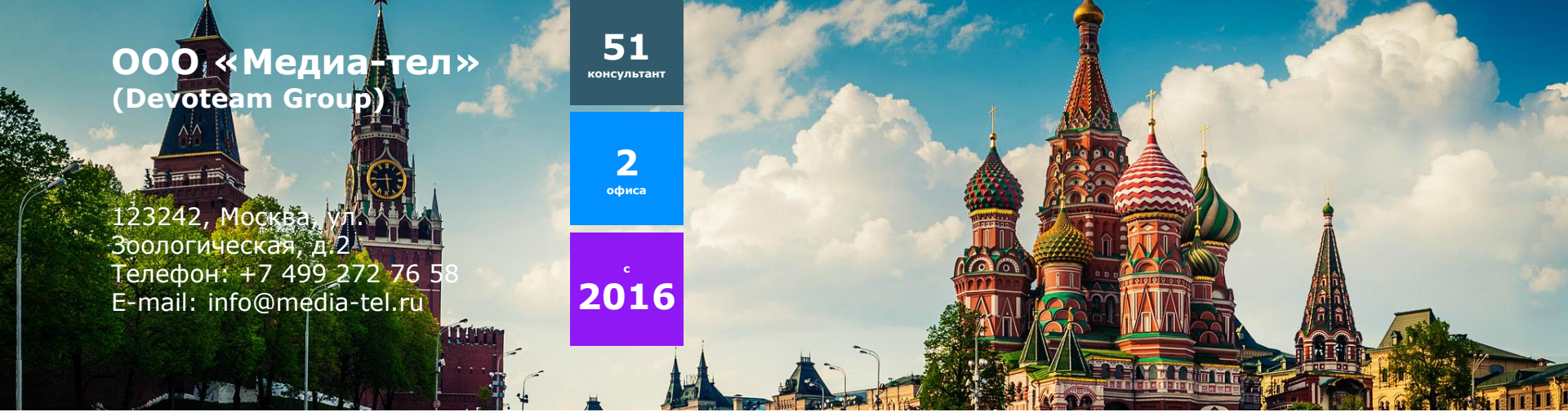
References

1. Process Mining Online Course: https://www.coursera.org/learn/process-mining/home/welcome
2. Leemans, Sander J.J.: Robust Process Mining with Guarantees
3. Sander J.J. Leemans: Inductive visual Miner & Directly Follows visual Miner
4. Wil M.P. van der Aalst: Process Mining in the Large: A Tutorial
5. Fabrizio Maria Maggi: Introduction to Process Discovery
6. Marc Solé Josep Carmona: An SMT-Based Discovery Algorithm for C-Nets
7. Wil Van der Aalst, RWTH Aachen University: A Practitioner's Guide to Process Mining: Limitations of the Directly-Follows Graph
8. Artem Polyvyanyy and Anna Kalenkova: Monotone Conformance Checking for Partially Matching Designed and Observed Processes

Contacts

Евгений Забродский
evgeniy.zabrodskiy@media-tel.ru

MEDIA-TEL
by devoteam



ООО «Медиа-тел»
(Devoteam Group)

123242, Москва, ул. Зоологическая, д.2
Телефон: +7 499 272 76 58
E-mail: info@media-tel.ru

51 консультант

2 офиса

с **2016**

Константин Левчин
konstantin.levchin@devoteam.com
Тел.: +7 985 767 79 42

Кирилл Кушнарев
kirill.kushnarev@devoteam.com
Тел.: +7 916 160 12 16

Татьяна Орлова
tatiana.orlova@devoteam.com
Тел.: +7 919 106 52 15

Innovate Technology Consulting for business