



Методы и применения *Process Mining*

Алексей Мицюк, Лаборатория ПОИС ФКН НИУ ВШЭ

Коротко

- Что такое Process Mining?
- Каковы применения (основные задачи)?
- Классификация алгоритмов
- Есть ли какие-то реальные примеры применения?
- Ссылки, ПО (с чего начать, куда копать)

Теория

Данные – нефть XXI века

.000101

За последние **10 минут** в мире
было сгенерировано примерно
столько же данных, сколько за все
время существования
человечества до 2003 года.



Источник: Wil van der Aalst

Событийные данные

Большая часть данных, генерируемых в мире – это событийные данные. Источники событийной информации – информационные системы в разных отраслях:

- Транспорт
- Банки
- ИТ
- Страхование
- Туризм
- Производство
- Медицина
- Образование
- Интернет вещей

Журналы событий (Event Logs)

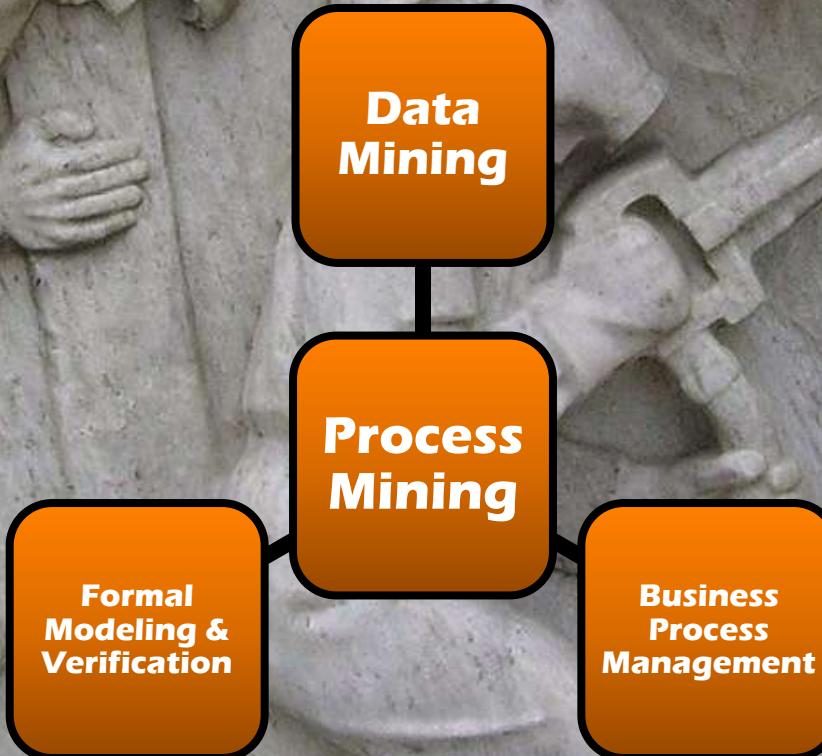
Слева: Снимок экрана Windows Event Viewer с логами событий.

Справа: Снимок экрана Windows PowerShell с выводом команды Get-EventLog -LogName application -Source Microsoft-Windows-Application-Activation-Service.

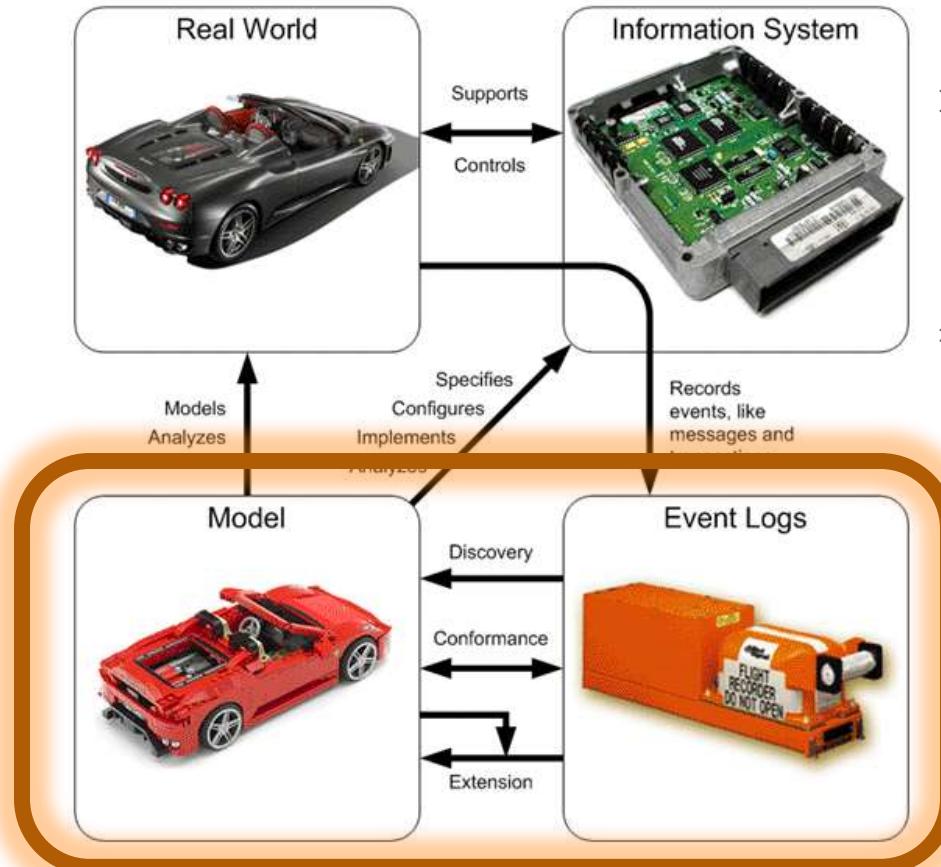
Внизу слева: Фото мужчины в шахтерской форме.

Внизу справа: Водяной знак на русском языке: "Что со всем этим делать?"

Process Mining



Process Mining – в реальном мире



Источник: processmining.org

Вопросы для Process Mining

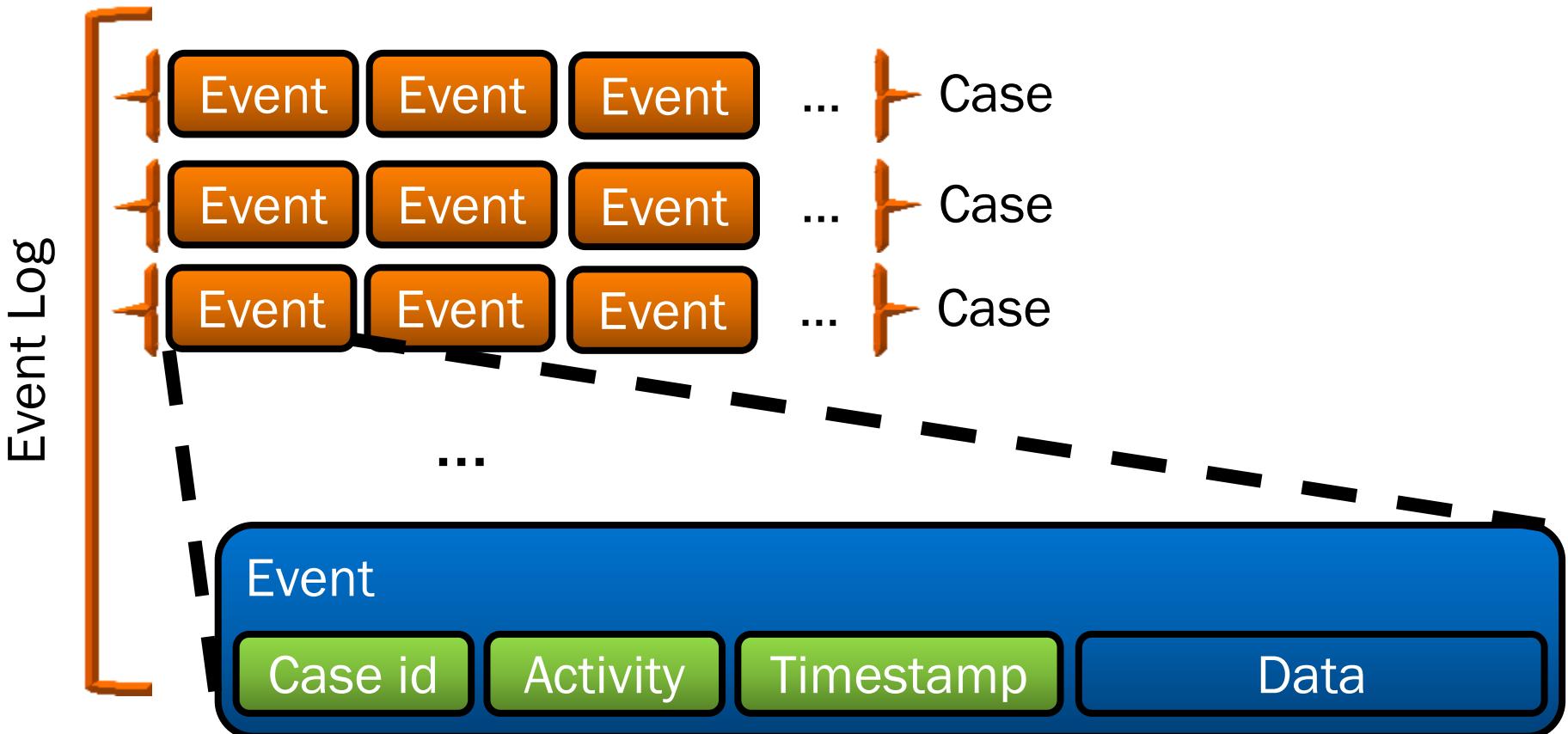
- Как на самом деле ведут себя люди (организации, устройства) в рамках процесса?
- Почему процесс исполняется не так эффективно, как предполагалось при проектировании?
- Где узкие места и тупики?
- В чем причины возникновения тупиков?
- Каковы типичные сценарии поведения?
- Чем реальное поведение участников процесса отличается от запроектированного?
- Как устроено взаимодействие участников процесса?
- Как изменения на одном (или не одном) шаге повлияют на процесс в целом?
- Как модифицировать процесс, чтобы повысить производительность?
- Когда (и где) высока вероятность появления проблем и узких мест?
- ...

Приложение в Software Engineering

Извлечение поведения пользователей

1. **Как ведут себя реальные пользователи?**
2. **Что НЕ получается у пользователей при работе с системой?**
3. **Какие функции нужны были бы, но отсутствуют?**
4. **Какие функции ПО НЕ используются, а какие более популярны?**
5. **Недостатки GUI – где пользователи попадают в тупик?**
6. **Особенности поведения для различных категорий пользователей, разных календарных периодов и т.д.**

Формат лога для Process Mining



Нотации моделирования

- Автоматы (Automata)
- Системы переходов (Transition Systems)
- Сети Петри (Petri Nets)
- Сети потоков работ (Workflow nets)
- BPMN (Business Process Model and Notation)
- BPEL (Business Process Execution Language)
- YAWL (Yet Another Workflow Language)
- EPC (Event-driven Process Chains)
- Исчисление процессов (Process Algebra)
- Деревья процессов (Process Trees)

Три раздела Process Mining

Process Discovery

- Лог событий ► Модель процесса

Conformance Checking

- Лог событий + Модель процесса ► Мера соответствия

Process Enhancement

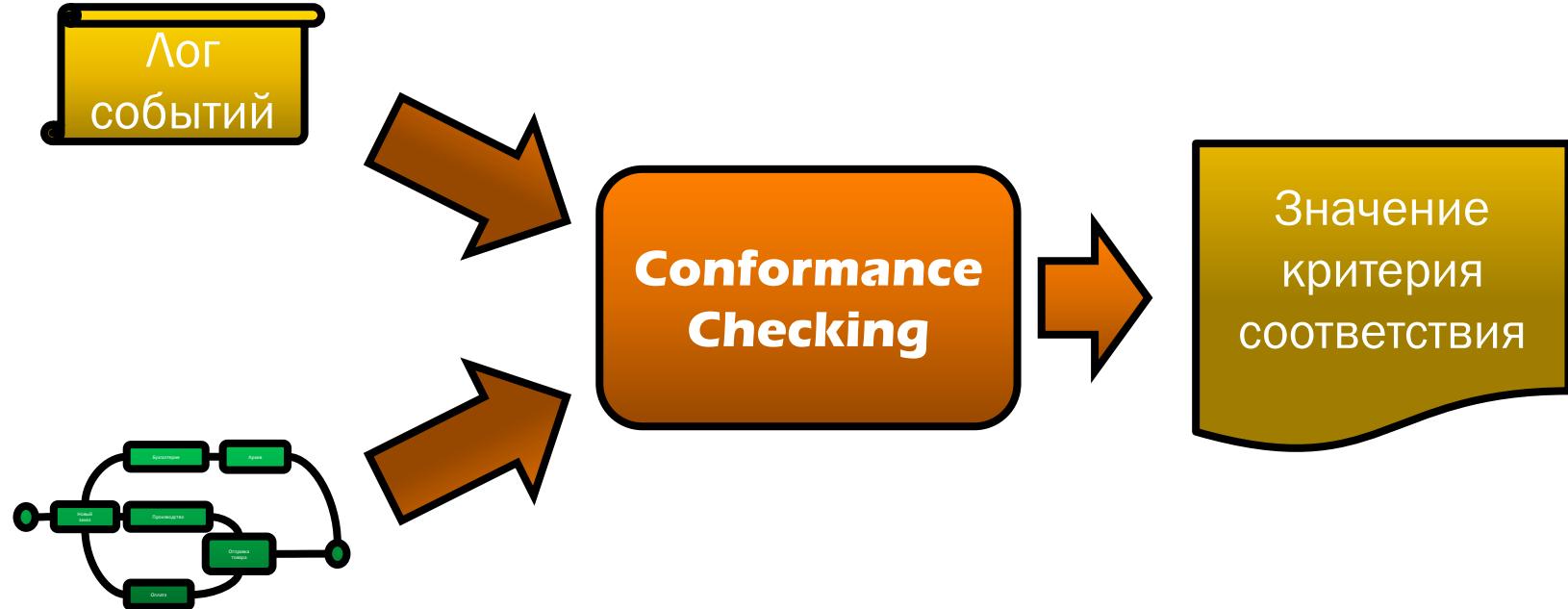
- Лог + Модель + Мера соответствия + ... ► Улучшенная модель

Три раздела Process Mining (1)



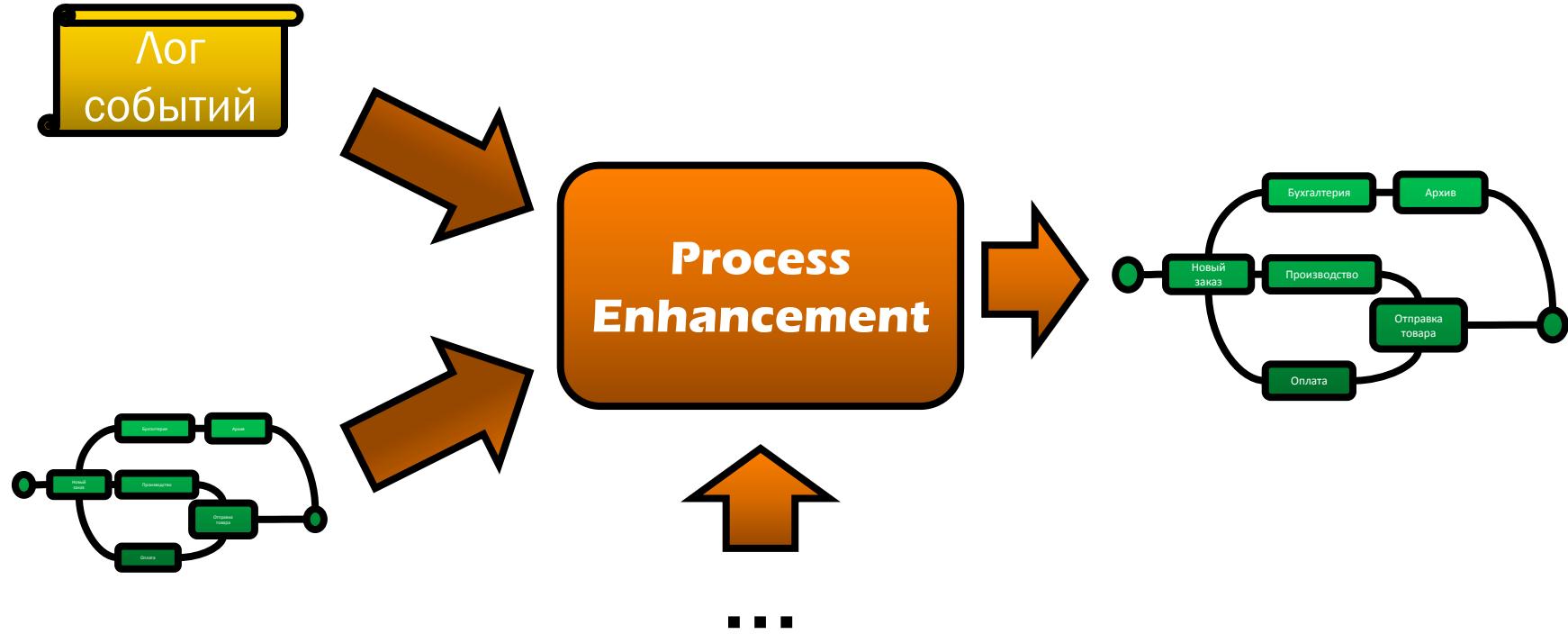
**Alpha Miner, ILP Miner, Heuristics Miner, Genetic Miner,
Fuzzy Miner, Inductive Miner, Region-Based Miners, ...**

Три раздела Process Mining (2)



Token-game Replay, Alignments Conformance Checker

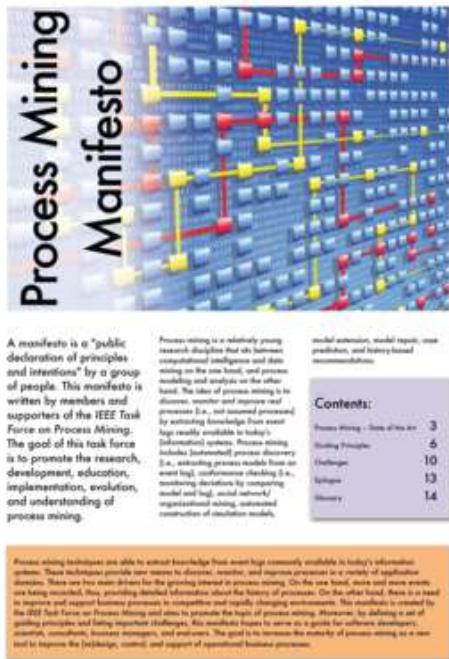
Три раздела Process Mining (3)



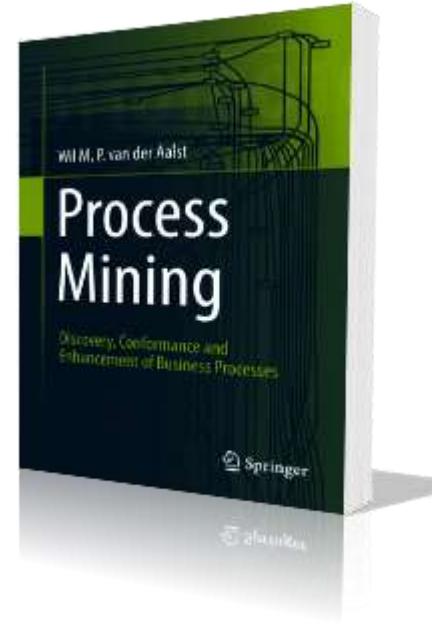
Три раздела Process Mining (4)



Ключевые материалы (1)



Process Mining Manifesto



Wil van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

Ключевые материалы (2)

<http://processmining.org/>

The screenshot shows the homepage of processmining.org. It features a top navigation bar with links for 'Search', 'Tools', 'Published', 'Event Log', 'Book', 'News', and 'Logout'. Below this is a main menu with 'Research' and 'Tools' sections. The 'Research' section includes a 'Process mining' logo and a brief description of what it is used for. The 'Tools' section includes a 'paism' logo and a brief description of its purpose.

<http://pais.hse.ru/>

The screenshot shows the homepage of the HSE PAIS Lab. It features a search bar at the top and a main content area with a large image and text about 'Научно-учебная лаборатория процессно-ориентированных информационных систем' (Scientific-educational laboratory of process-oriented information systems). Below this is a search interface with fields for 'Search', 'Type', and 'Category'.

<http://www.win.tue.nl/ieeetfpm/>

The screenshot shows the homepage of the IEEE CIS Task Force on Process Mining. It features a yellow header with the task force's name and a main content area with a search bar and a list of items under 'Recent news'.



ФУНКЦИОНАЛ

RapidProM



ProM Framework

PMLAB

VTMine/C

Инструменты

perceptive software
from Lexmark

**Perceptive
Process Mining**

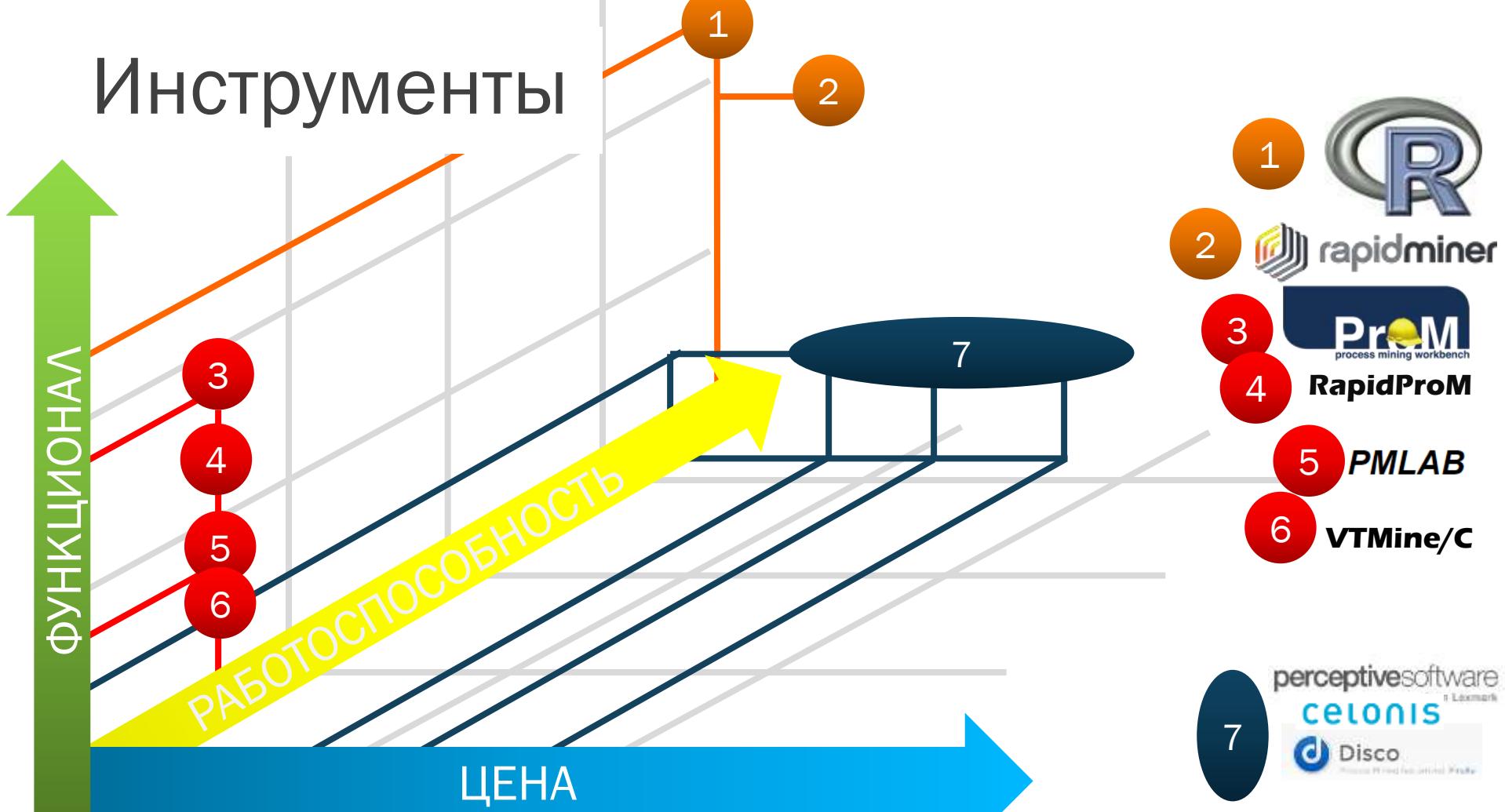
celonis
Celonis
Discovery

Fluxicon Disco



ЦЕНА

Инструменты



Инструменты: коммерческие

- Fluxicon Disco
- Perceptive Process Mining (before Futura Reflect and BPM|one)
(Perceptive Software)
- ARIS Process Performance Manager (Software AG)
- QPR ProcessAnalyzer (QPR)
- Celonis Discovery (Celonis)
- Interstage Process Discovery (Fujitsu)
- Discovery Analyst (StereoLOGIC)
- XMAnalyzer (XMPro)
- и другие
- Отдельные алгоритмы в составе инструментов BI от IBM, SAP

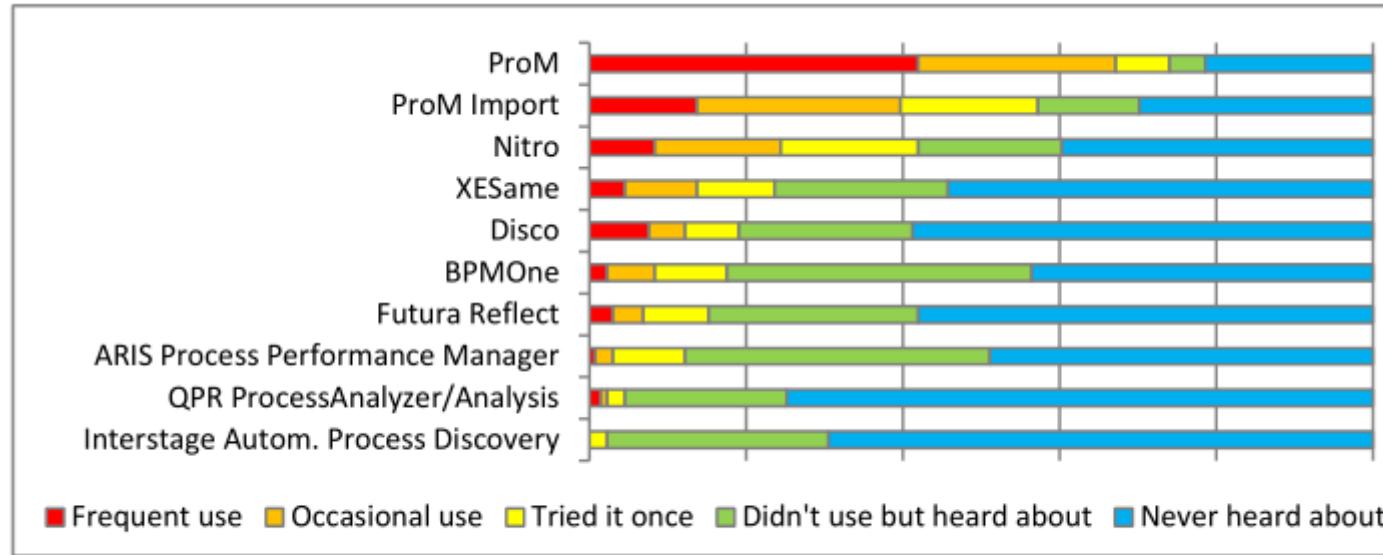


Fig. 4. Tools for process mining (question 2², 119 respondents)



- Open source
(Лицензия GPL)
- Java-based
- Пользовательские
расширения
- Формат XES



<http://www.promtools.org/>

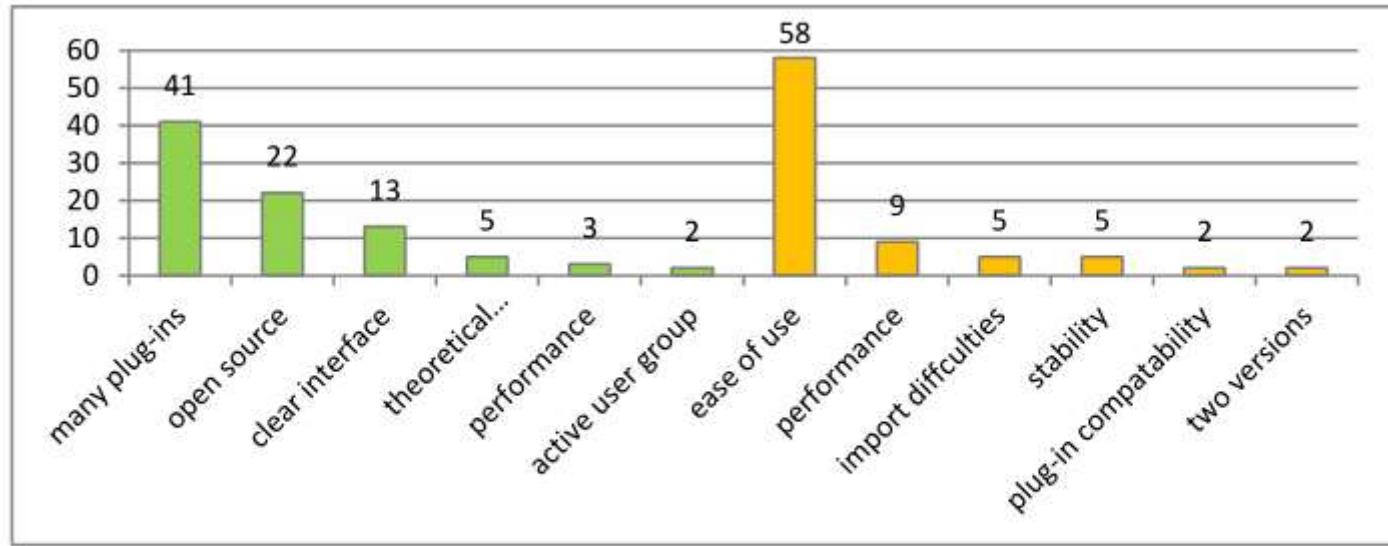


Fig. 5. Benefits and drawbacks of ProM (question 9², 78 respondents)
(green: benefits, orange: drawbacks)



<http://www.xes-standard.org/>

```
        <string key="concept:name" value="Run tests"/>
        <string key="concept:instance" value="Run tests"/>
    </event>
    <event>
        <string key="concept:name" value="End iteration"/>
        <string key="concept:instance" value="End iteration"/>
    </event>
</trace>
<trace>
    <string key="concept:name" value="Trace 3"/>
    <event>
        <string key="concept:name" value="Start iteration"/>
        <string key="concept:instance" value="Start iteration"/>
    </event>
    <event>
        <string key="concept:name" value="Learn requirements"/>
        <string key="concept:instance" value="Learn requirements"/>
    </event>
    <event>
        <string key="concept:name" value="Develop tests"/>
        <string key="concept:instance" value="Develop tests"/>
    </event>
    <event>
        <string key="concept:name" value="Code"/>
    </event>

```



Anonymous log imported from new_log_1.xes

Create new...



Dashboard



Inspector



Summary

Key data

Processes 1

Cases 10

Events 125

Event classes 7

Event types 1

Originators 1

Events per case



Event classes per case



Log info

Start date

End date

Anonymous log imported from new_log_1.xes

Create new...



Log inspector

Browser

Explorer

Log Attributes



Dashboard



Inspector



Summary

Instances

- Trace 1
- Trace 2
- Trace 3
- Trace 4
- Trace 5
- Trace 6
- Trace 7
- Trace 8
- Trace 9**
- Trace 10

Trace 9
6 events

Start iteration
#1
- no timestamp -

Learn requirements
#2
- no timestamp -

Develop tests
#3
- no timestamp -

Code
#4
- no timestamp -

Run tests
#5
- no timestamp -

End iteration
#6
- no timestamp -

Attributes for case Trace 9

conceptname: Trace 9

Actions

Activity...



Input

Anonymous log imported from new_lo...
XLog

Click to add input object

Actions

Filter:



alp



Mine for a Petri Net using Alpha-algorithm

B.F. van Dongen (b.f.v.dongen@tue.nl)

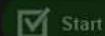
AlphaMiner

Output

Click to add output object



Reset



Start



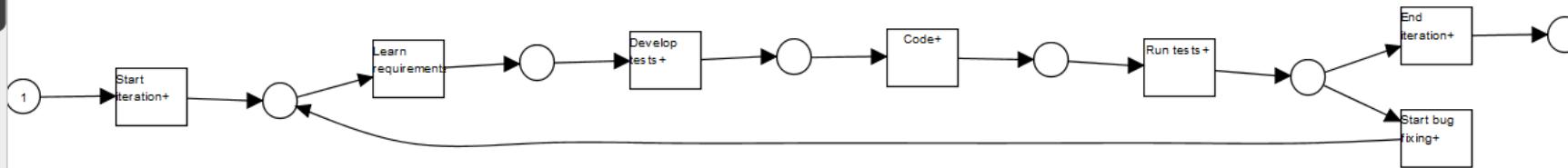
Petri net from Anonymous log imported from new_log_1.xes , mined with AlphaMiner

Create new...



PIP

Zoom



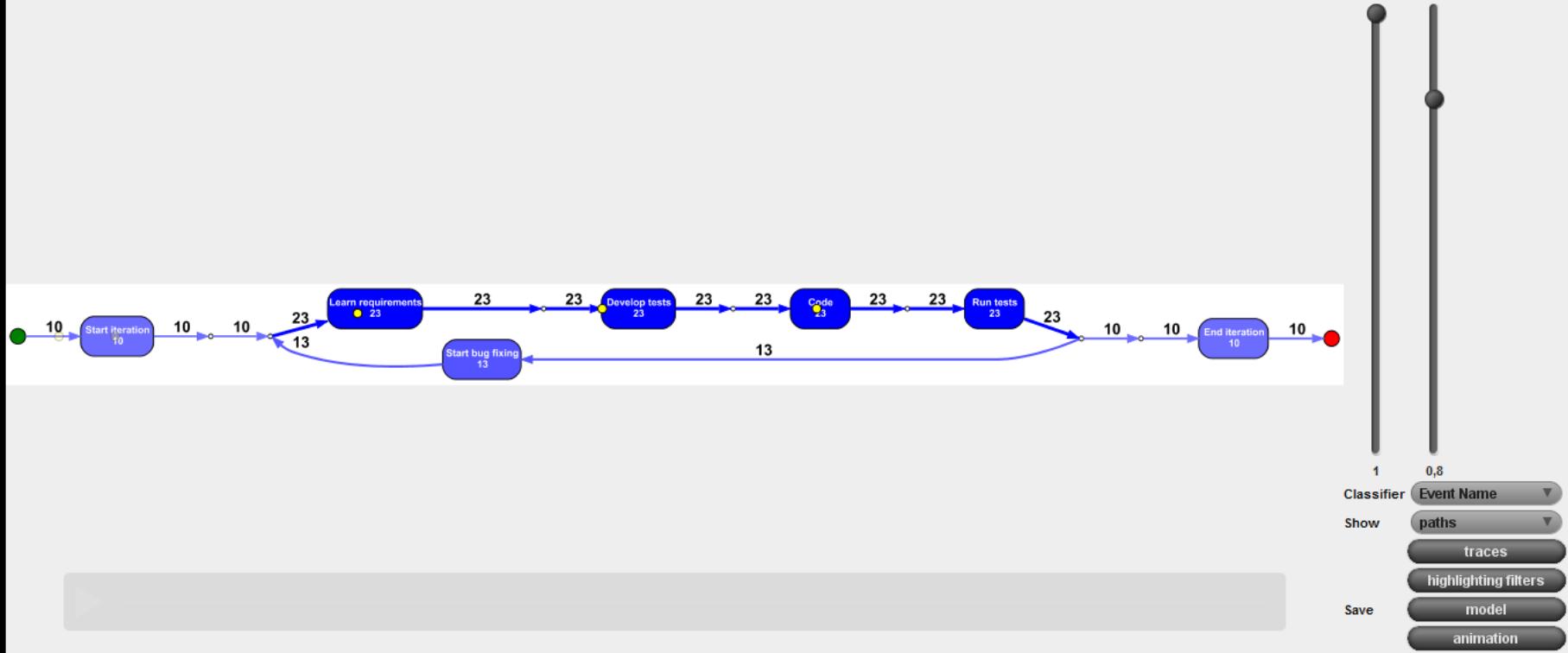
Export

Anonymous log imported from new_log_1.xes

Create new...



activiti... paths:



Практика

Примеры применения (2011-2014)

0	Monitoring Deployed Application Usage with Process Mining	C.W. Günther, A. Rozinat, W.M.P. van der Aalst, Kenny van Uden
1	Process Mining of Test Processes: A Case Study	A. Rozinat, I.S.M. de Jong, C.W. Günther, and W.M.P. van der Aalst
2	Process Mining of Event Logs in Auditing: A Field Study of Procurement at a Global Bank (2012)	M. Jans, M. Alles & M. Vasarhelyi
3	Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital	R.S. Mans, M.H. Schonenberg, M. Song, W.M.P. van der Aalst, and P.J.M. Bakker
4	Configurable Process Models: Experiences from a Municipality Case Study	Florian Gottschalk, Teun A.C. Wagelmakers, Monique H. Jansen-Vullers, Wil M.P. van der Aalst, Marcello La Rosa
5	Understanding Process Behaviours in a Large Insurance Company in Australia: A Case Study	S. Suriadi, M.T. Wynn, C. Ouyang, A.H.M. ter Hofstede, N.J. van Dijk
6	Using process mining for the analysis of an e-trade system: A case study	A. Mitsyuk, A. Kalenkova, S. A. Shershakov, W.M.P. van der Aalst

Пример применения: исходные данные

Российский портал по продаже электронных авиа-билетов

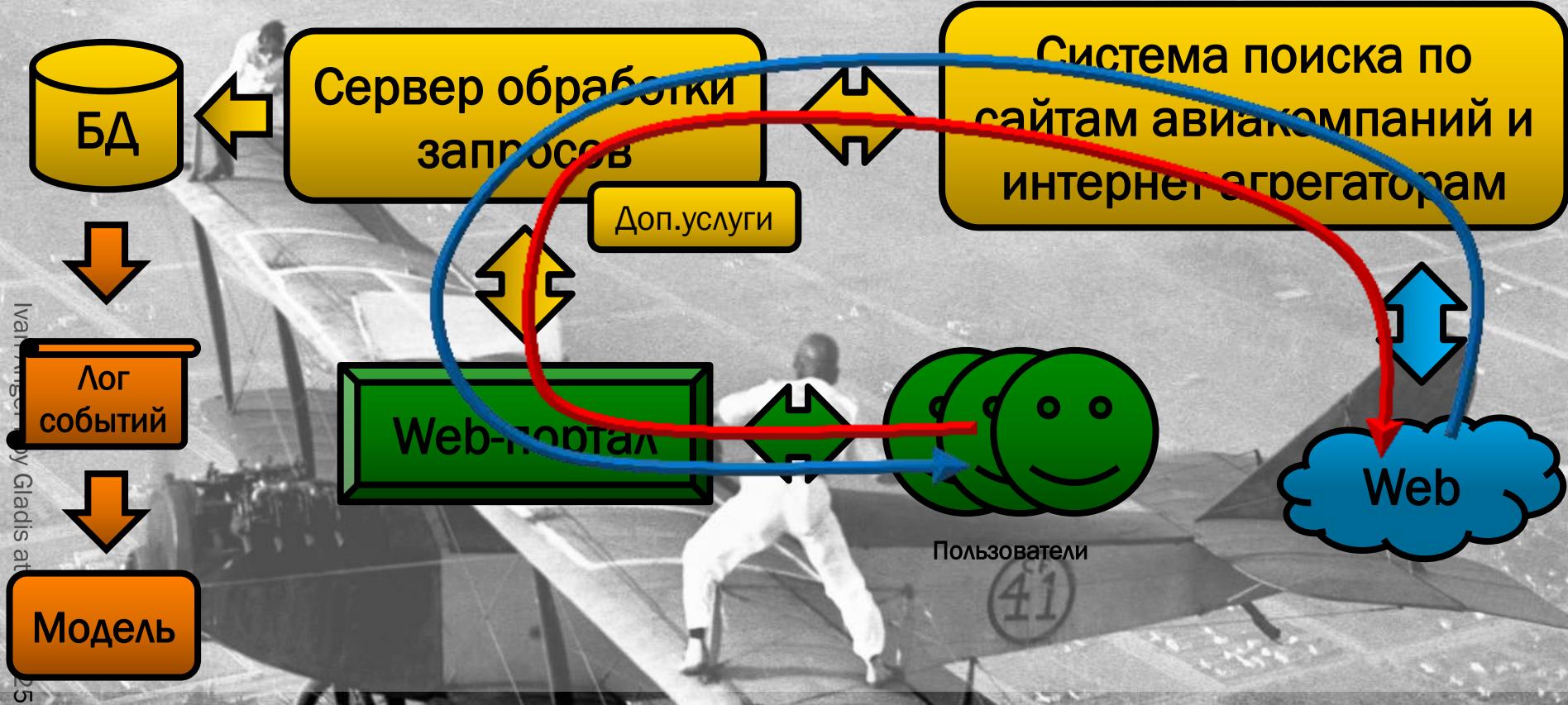
Web-портал помогает искать предложения от разных авиакомпаний

Позволяет осуществлять бронирование разных авиа-билетов, дополнительное страхование

Пользователи взаимодействуют с системой через сайт

Поведение пользователей на сайте записывается в базу данных, содержащую несколько таблиц, каждая из которых содержит определенную информацию

Схема информационной системы



Задача

Согласно исследованию, средняя доля совершивших покупку из посетителей сайта, ниже, чем средняя по отрасли.

Почему?

Найти узкие места

Препроцессинг



Лога событий в необходимом формате не было.

Для получения лога были использованы запросы SQL.
идентификатор сессии >> Идентификатор экземпляра процесса

По идентификатору сессии из таблиц выбирались соответствующие записи.
Действие << пара записей «действие пользователя» (click, scroll, select и т.д.) и
«объект действия» (submit button, select booking button и т.д.)

Временные метки в БД, к счастью, сохранялись

Session ID	Activity
1054514	OW-LO
	LE-CHN
	PT-CHN
	SUBMIT
	W-UNL

Анализ проводился на данных за
(примерно) месяц работы портала

Таким образом удалось получить журнал
событий, содержащий все необходимое
для Process Mining, а также кучу
дополнительных данных (код заказа,
направление, ОС заказчика и т.д.)

Анализ журнала событий

Для анализа использовался ProM 6 Framework



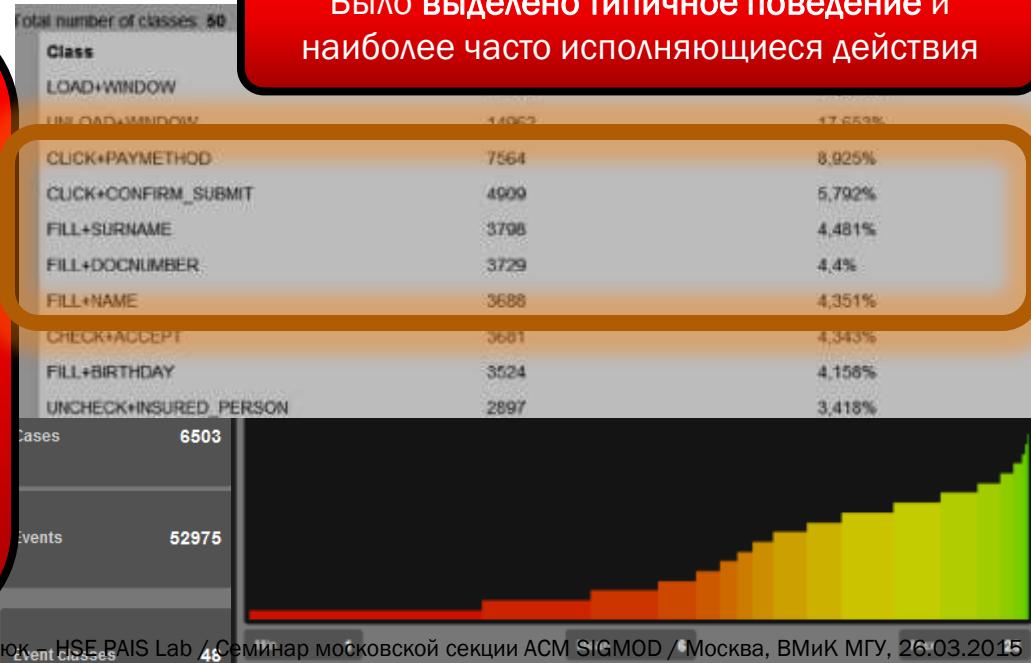
Прежде всего провели анализ лога

Очень удобно, что в нем содержались события загрузки и выгрузки страницы, отмечающие начало и окончание сессии.

Оказалось, что есть **куча** (около 40%) случаев, когда страница загрузилась, а потом выгрузилась через пол минуты.

Работы были отфильтрованы!

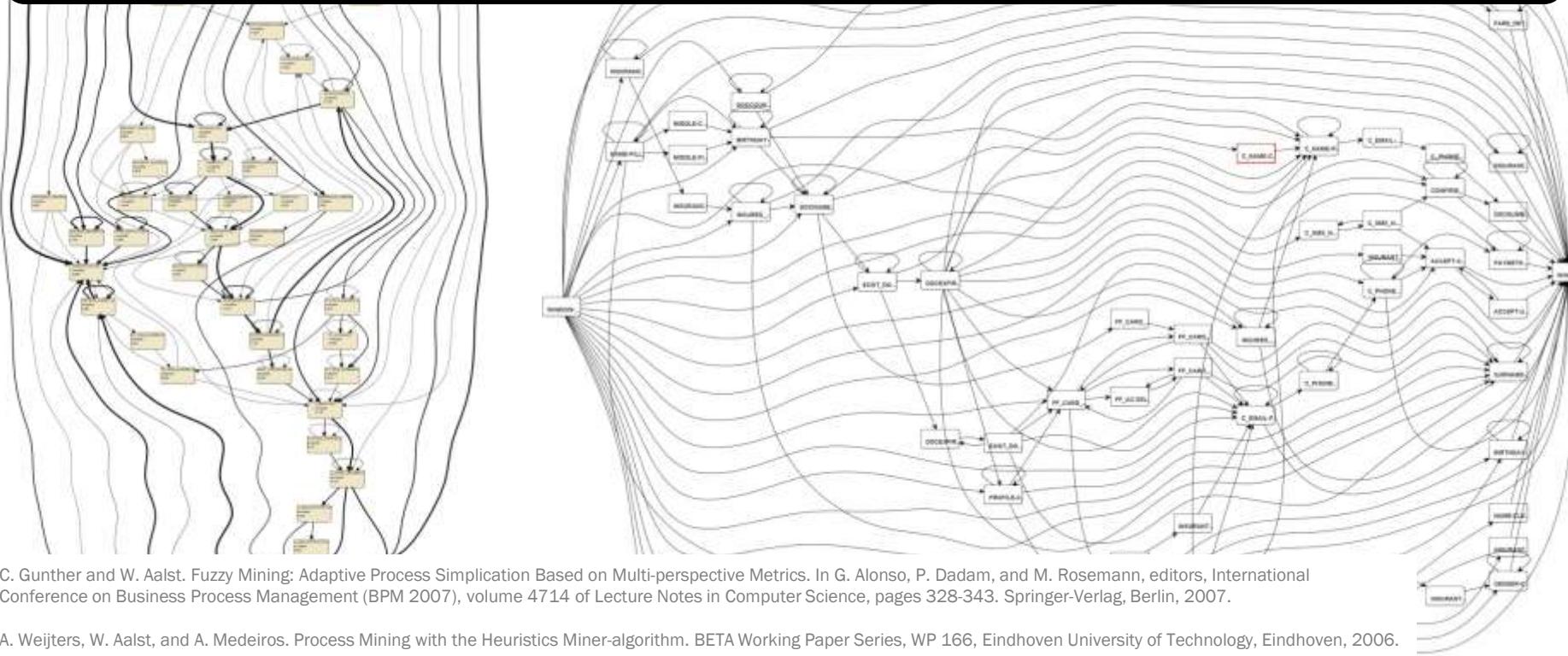
Было выделено типичное поведение и наиболее часто исполняющиеся действия





Нечеткая модель

Модель поведения пользователей (без роботов) построена с помощью Fuzzy и Heuristics Miner

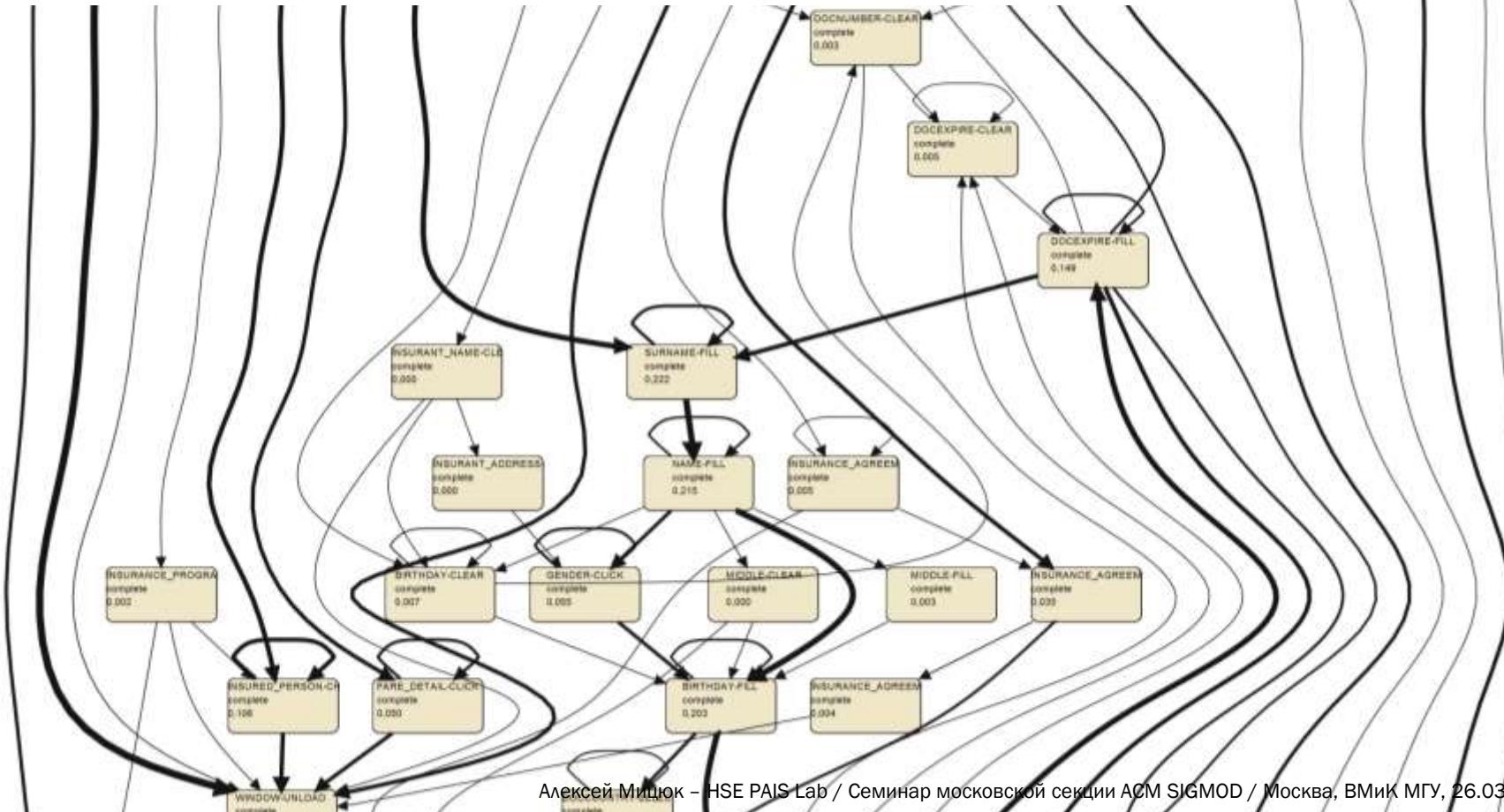


C. Gunther and W. Aalst. Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics. In G. Alonso, P. Dadam, and M. Rosemann, editors, International Conference on Business Process Management (BPM 2007), volume 4714 of Lecture Notes in Computer Science, pages 328-343. Springer-Verlag, Berlin, 2007.

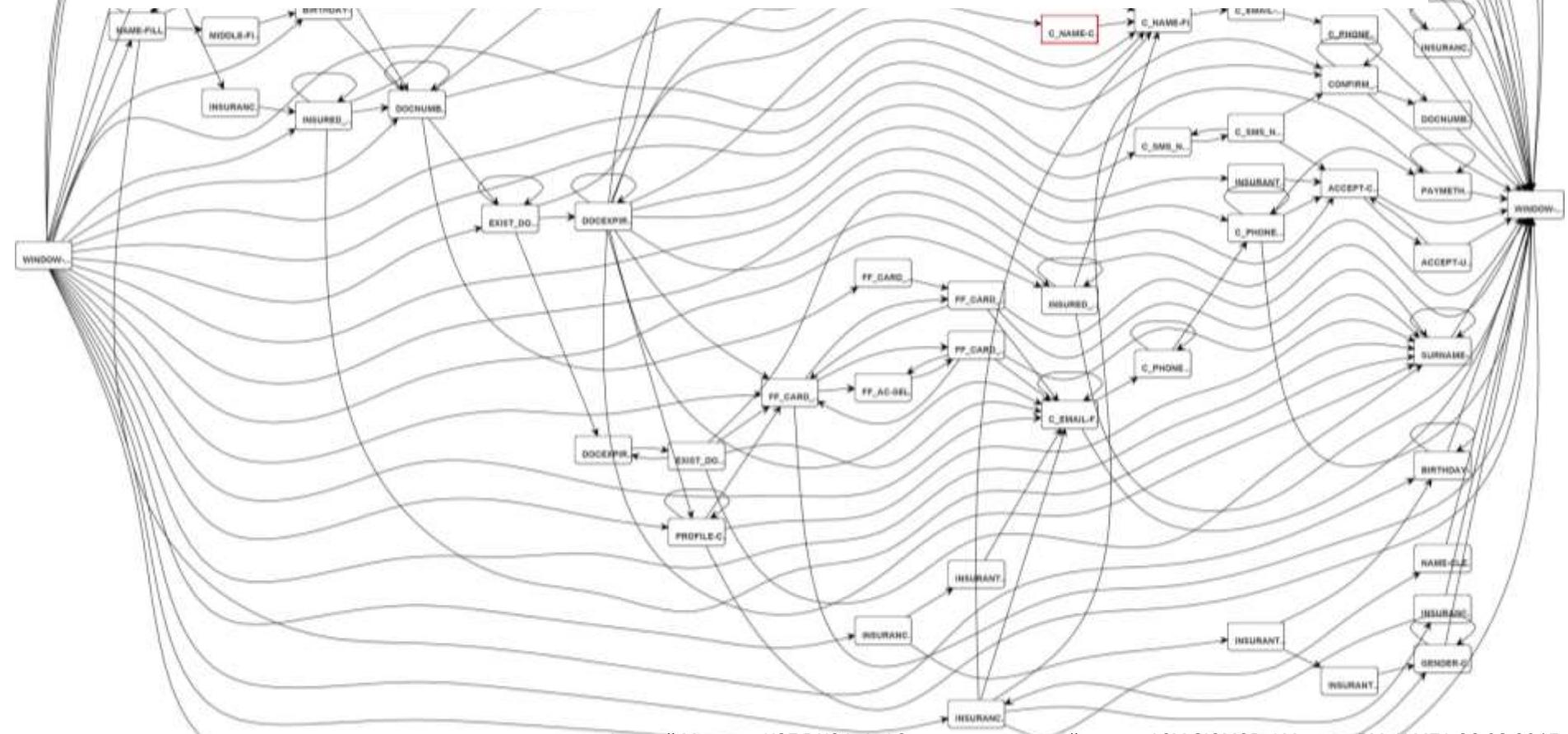
A. Weijters, W. Aalst, and A. Medeiros. Process Mining with the Heuristics Miner-algorithm. BETA Working Paper Series, WP 166, Eindhoven University of Technology, Eindhoven, 2006.

Нечеткая модель (Fuzzy Miner)

MOVE_NUMBER-CLEAR
complete
0.003



Эвристическая сеть (Heuristics Miner)



Выводы (1)

Что показывают эти жуткие модели?

- Последовательность действий пользователей (все варианты)
- Пользователи ведут себя по разному даже на линейно выстроенной странице
- Тем не менее, более половины ведут себя довольно единообразно (ввод даты и направления, поиск предложения, просмотр условий, ввод личных данных, выбор дополнительных опций, ввод платежных данных, покупка)
- Есть люди, которые платят без поиска предложения. Удивительно? Это **тестировщики!**
- Логи в день тестирования вообще отличаются от обычных: оказывается, тестировщики думают, что пользователи ведут себя не так, как они себя ведут😊
- Есть **около 20 случаев странного поведения** (типа поиска десятков предложений или многократных попыток ввода различных платежных данных)

Выводы (2)

Что пользователи делают чаще всего перед тем, как покинуть портал?

(1) Пользователи покидают портал, когда **пытаются отправить свои данные**. Почему? Оказывается, есть довольно много трасс в логе, которые завершаются многократным повторением события CONFIRM_SUBMIT. Проблема решается **расширением канала** от web-портала к серверу, а также **устранением ошибок в ПО**.

Вообще, **многие сценарии** взаимодействия с сайтом были **не предусмотрены** разработчиками. Некоторые из элементов управления не работали в той последовательности, в которой их пытались использовать.

(2) Пользователи покидают портал, когда **узнают условия оплаты**. Почему? Оказалось, что недавно **компания изменила условия продаж**, а именно возврата средств при отмене заказа для некоторых типов билетов. Руководство не могло предположить, что усложнение порядка возврата так изменит картину покупок.

(3) Пользователи покидают портал, когда **выбирают способы оплаты**. Почему? Оказывается, портал испытывает **проблемы с оплатой картами** некоторых распространенных **платежных систем**, а также не принимает оплату через некоторые системы интернет-банкинга. Записи об этом уже даже есть в «жалобной книге» портала (без внимания службы поддержки, конечно же)!

(4) Пользователи покидают портал, когда видят, что им **предлагают дополнительно застраховаться**. Почему? Здесь проблема **неправильного проектирования GUI**. Сначала пользователю показывался экран, где была отмечена доп. страховка, а снять её он мог только на следующем экране. Многие просто закрывали сайт.

Выводы (3)

Были предложены пути по исправлению ситуации

- Переработка GUI с изменением способа и порядка отображения предложений
- Изменение способа показа пользователю и **выбора дополнительных услуг**
- Исправление ошибок в ПО backend'а и повышение скорости отклика
- Усовершенствование схемы покупки
- Пересмотр тарифов и политики возврата билетов
- Расширение спектра поддерживаемых систем оплаты



Почему нечеткие и
эвристические модели?

- Потому что шум!

Почему модель слабо
структурирована («спагетти»)?

- Потому что поведение живых
людей!

Почему без BPMN?

- Потому что BPMN-майнинг
появился позже. Теперь
можно и в BPMN!

Почему не Data-Mining?

- И он тоже, нет противоречия!

Почему не полная автоматика
на лету (on-the-fly)?

- А действительно, почему бы
и нет? Разработки ведутся!

Главный вывод

Применение Process Mining для анализа поведения пользователей информационных систем зависит прежде всего от **наличия подходящих логов и от разумной постановки вопросов**

Литература

- IEEE CIS Task Force on Process Mining. Process Mining Manifesto. LNBI 99, pp. 169-194. Springer, 2012.
DOI: 10.1007/978-3-642-28108-2_19
- Wil van der Aalst. Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer, 2011.
- Marlon Dumas, Wil M. van der Aalst, Arthur H. ter Hofstede. Process-Aware Information Systems: Bridging People and Software Through Process Technology. Wiley, 2005. ISBN: 978-0-471-66306-5
- Wil M. P. van der Aalst. Business Process Management: A Comprehensive Survey // ISRN Software Engineering, Volume 2013 (2013), Article ID 507984.
- Alexey Mitsyuk, Anna Kalenkova, Sergey A. Shershakov, van der Aalst W. Using process mining for the analysis of an e-trade system: A case study // Бизнес-информатика. 2014. Vol. 29. No. 3. P. 15-27.
- R.S. Mans, W.M.P. van der Aalst, R.J.B. Vanwersch, A.J. Moleman. Process Mining in Healthcare: Data Challenges when Answering Frequently Posed Questions. In ProHealth 2012/KR4HC 2012 Proceedings. Springer, 2013.
- Suriadi et al. Understanding Process Behaviours in a Large Insurance Company in Australia: A Case Study. In Advanced Information Systems Engineering - LNCS, pp. 449-464. Springer, Valencia, Spain, 2013.
- Научно-учебная лаборатория процессно-ориентированных информационных систем (ПОИС). -- <http://pais.hse.ru/>
- Process Mining: Research, Tools, Applications. -- <http://www.processmining.org/>
- Process Mining Workbench. -- <http://www.promtools.org/>



Монумент **Arbetets ära och glädje**

(Честь, труд и радость) в городе

Eskilstuna (Швеция)

Автор: **Ivar Johnsson (1942)**

Фото: wikipedia (by **Raphael Saulus**)

Спасибо за внимание!

<http://pais.hse.ru>