

# Process Mining своими руками

с кратким введением в историю вопроса  
примерами и извинениями.

Дмитрий Колодезев, ООО Промсофт, Новосибирск  
[dk.promsoft@mail.ru](mailto:dk.promsoft@mail.ru)

# Краткое содержание

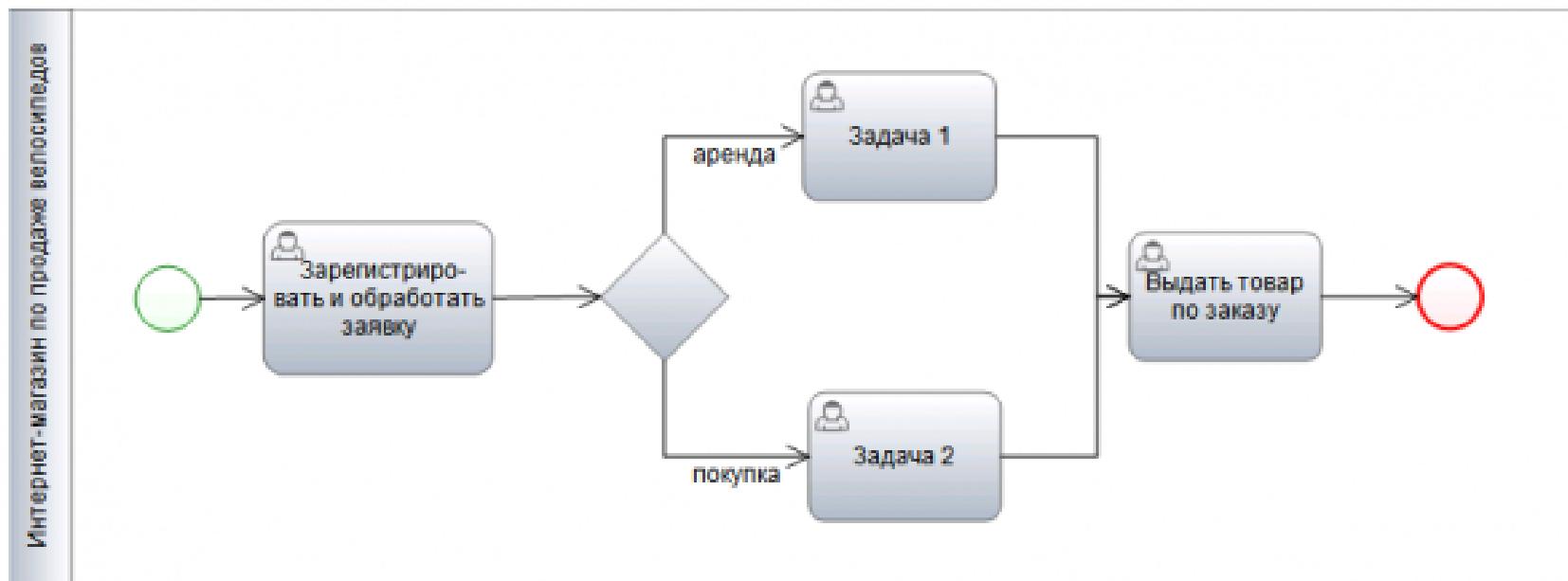
- РМ — изучение процессов на основе логов
- Задачи:
  - синтез моделей
  - проверка соответствия
  - анализ процессов
- Есть теория
- Есть специализированный софт
- Простые вещи делаются на коленке
- Есть куда копать

# Краткое содержание - 2

1. Что такое Process Mining
2. История вопроса
3. Процессы, более строго.
4. Логи
5. Параллельные элементы
6. Альфа Алгоритм
7. Специализированный софт
8. Разведочный анализ своими силами, на игрушечном примере.
9. Аннотированный список книг, курсов, ссылок.
10. За пределами WFMS логов.
11. Процессы как текст.
12. Вопросы

# Что такое РМ - 1

- РМ — изучение процессов на основе логов
- "Сколько еще лет мы будем моделировать процессы через интервью участников?"



Такие картинки рисовали во всех бизнес-школах, начиная с 1960-х

# Что такое РМ - 2

- Госпиталь:
  - Почему больные так долго ждут операции?
  - Следуют ли врачи регламенту?
  - Можем ли мы предсказать время ожидания?
  - Сколько людей нужно и где?
  - Как снизить затраты (времени, денег)

Понять и улучшить поток оказания услуги в госпитале, проанализировав записи, разбросанные по сотням таблиц базы данных.

# Что такое РМ - 3

- Philips, томографы:
  - Как на самом деле используются машины?
  - Почему они ломаются?
  - Какие компоненты нужно заменить?
  - Можем ли мы предсказать поломку?
  - Какие части нужно улучшить?
  - Каким должен быть тестовый сценарий?

Понять и улучшить работу и характеристики томографов, используя данные, полученные от сети сенсоров.

# Что такое РМ - 4

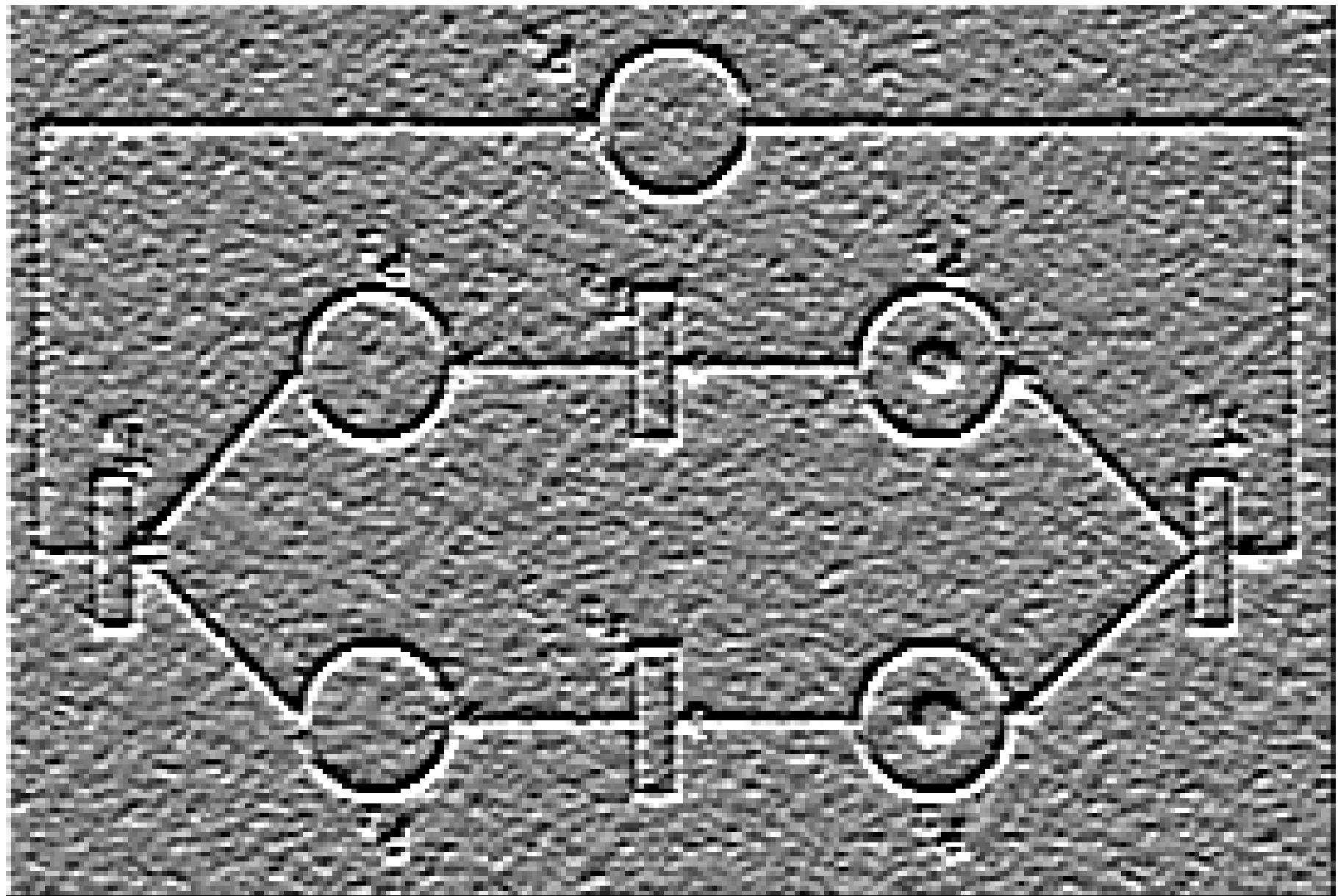
- Каков процесс работы на самом деле?
- Где узкие места?
- Где люди и машины отклоняются от ожидаемых или предписанных процедур?
- Какие участки работают хорошо?
- Какие факторы влияют на узкие места?
- Можем ли мы предсказать задержки, отклонения, риски для уже идущих процессов?
- Что мы можем предпринять, чтобы снизить риски/цены etc?
- Как нам преобразовать Рабкрин?

Возражение это состоит в том, что будто бы из предлагаемого мной преобразования получится один хаос. Члены ЦКК будут слоняться по всем учреждениям, не зная, куда, зачем и к кому им обратиться, внося повсюду дезорганизацию, отрывая служащих от их текущей работы, и т. д. и т. п. (с) В.И. Ленин

# Что такое РМ - 5

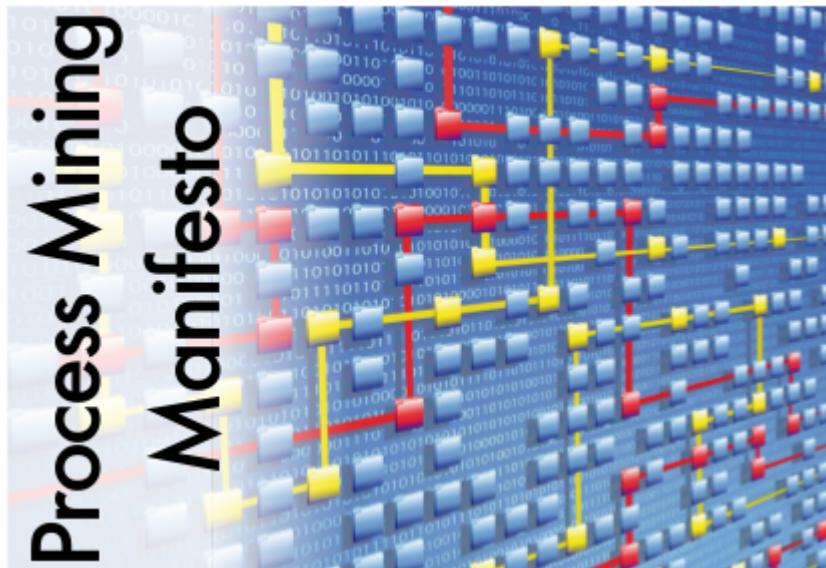
- Процессы нужно изучать в динамике
- Псевдо динамика. Множественные замеры.  
Фото vs видео. Недостатки.
- Частотный спектр vs временной ряд
- Информационные системы пишут логи
- Базовые задачи:
  - Синтез модели
  - Проверка соответствия
  - Совершенствование процессов
  - Визуализация

# История вопроса



# История вопроса - 2

- 2011 BPM 2011 International Workshops



Process mining techniques are able to extract knowledge from event logs commonly available in today's information systems.

These techniques provide new means to discover, monitor, and improve processes in a variety of application Domains.

.. this manifesto hopes to serve as a guide for software developers, scientists, consultants, business managers, and end-users.

IEEE CIS Task Force on Process Mining

Eindhoven University of Technology TU/e

ВШЭ

# История вопроса - 3



Это всё он придумал

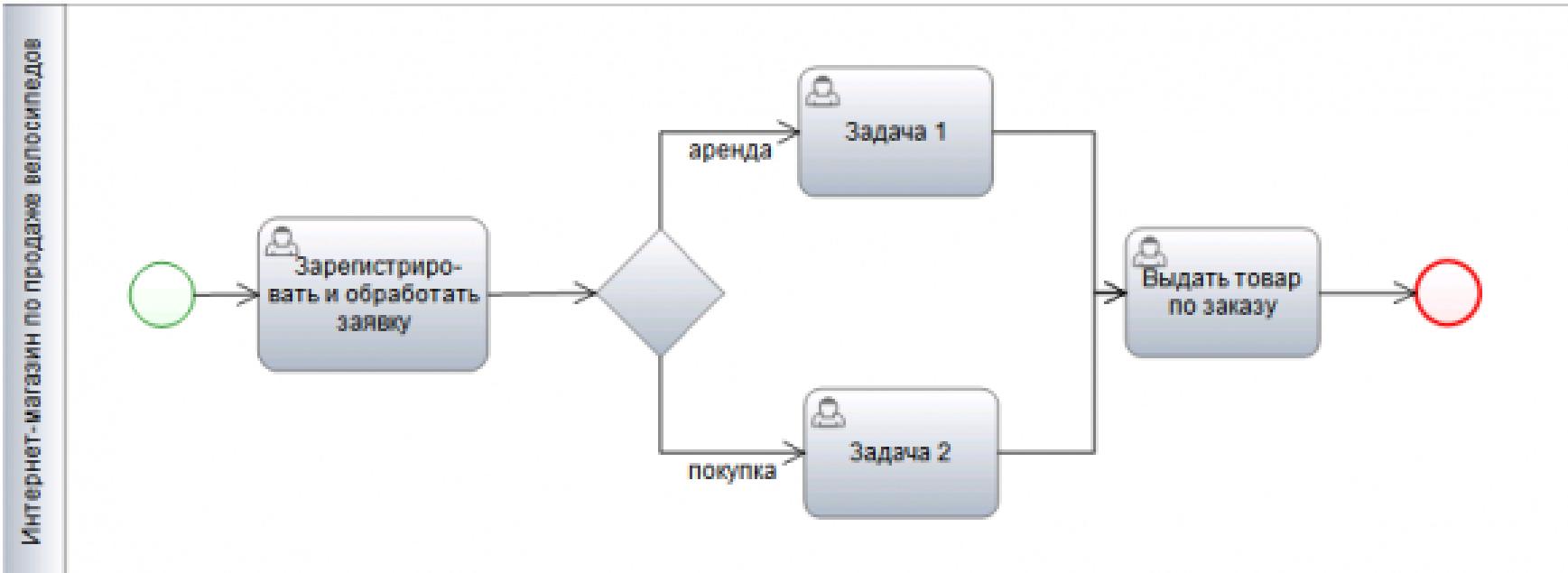
Wil van der Aalst, TU/e

# Процесс

## ПРОЦÉСС

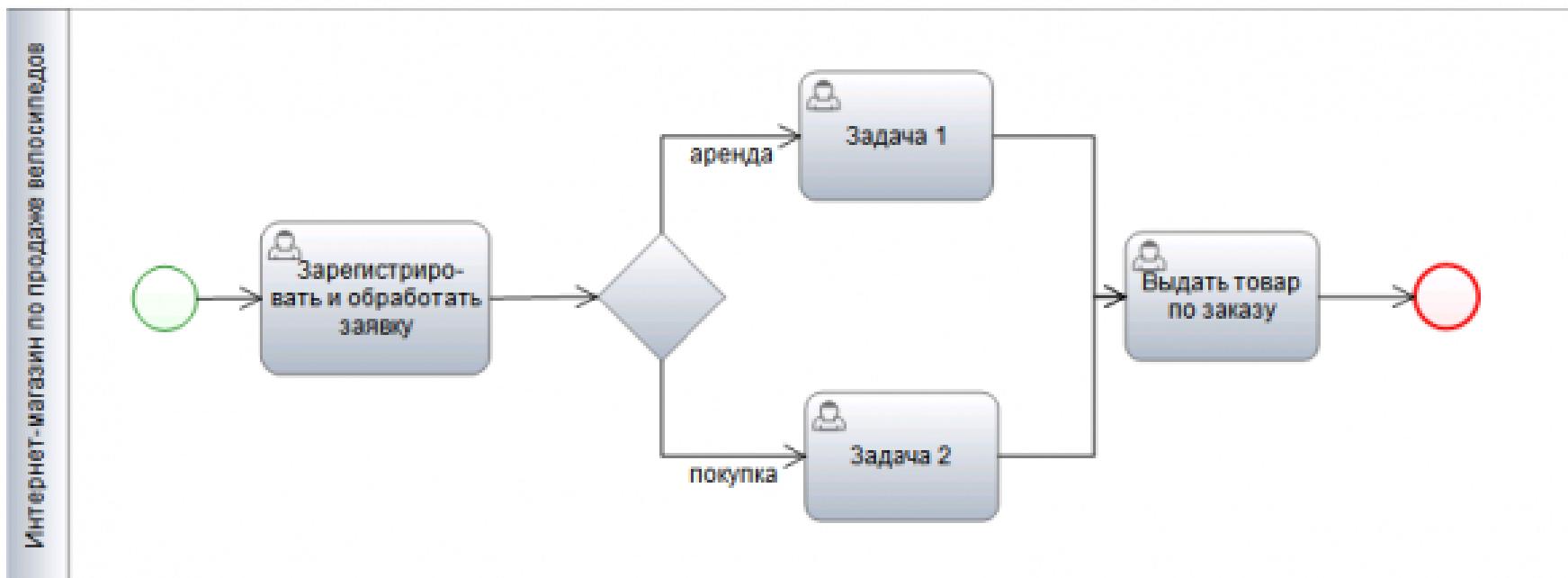
Мужской род

1. Ход, развитие какого-н. явления, последовательная смена состояний в развитии чего-н.  
"П. работы. В процессе борьбы"
2. Порядок разбирательства судебных дел, а также судебное дело.  
"Уголовный п."



# Процесс - 2

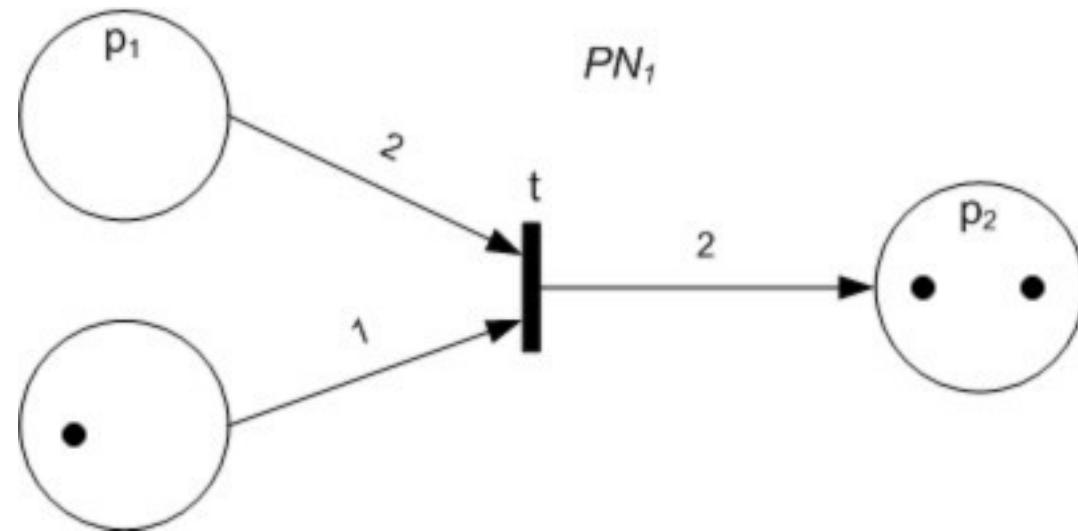
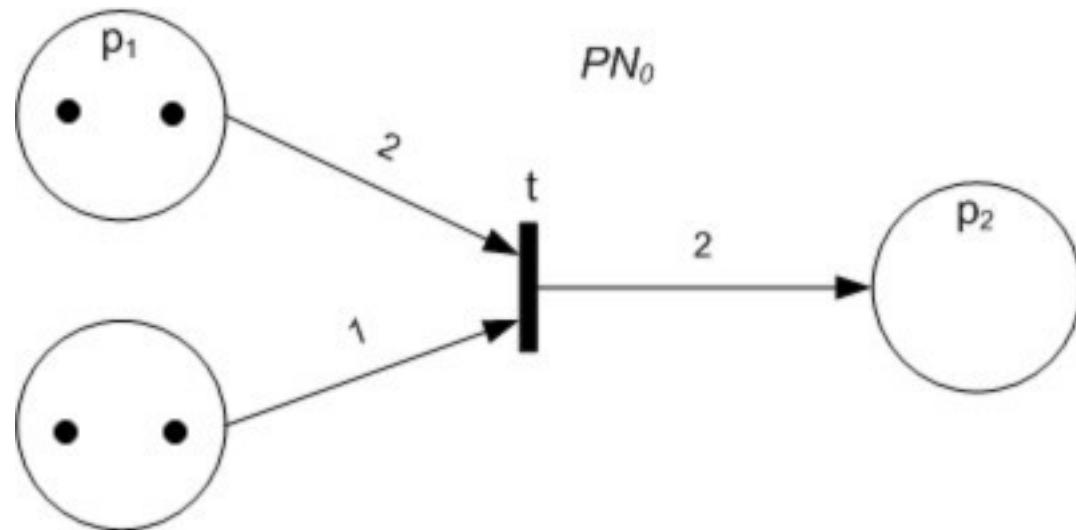
- Процесс
- Экземпляр процесса (case)
- След процесса (trace)
- Событие (event)
- Ресурсы



# Petri net

- Карл Адам Петри, 1939, химические процессы
- $N = (P, T, F, W, m^*)$
- места, переходы, дуги, веса, разметка
- $P \cap T = \emptyset$
- $F \subseteq (P \times T) \cup (T \times P)$
- $W : (P \times T) \cup (T \times P) \rightarrow N$
- $W(x, y) > 0$  iff  $(x, y) \in F$
- $m^* : P \rightarrow N$
- Временная, раскрашенная, стохастическая, ингибиторная, иерархическая ....

# Petri net - 2



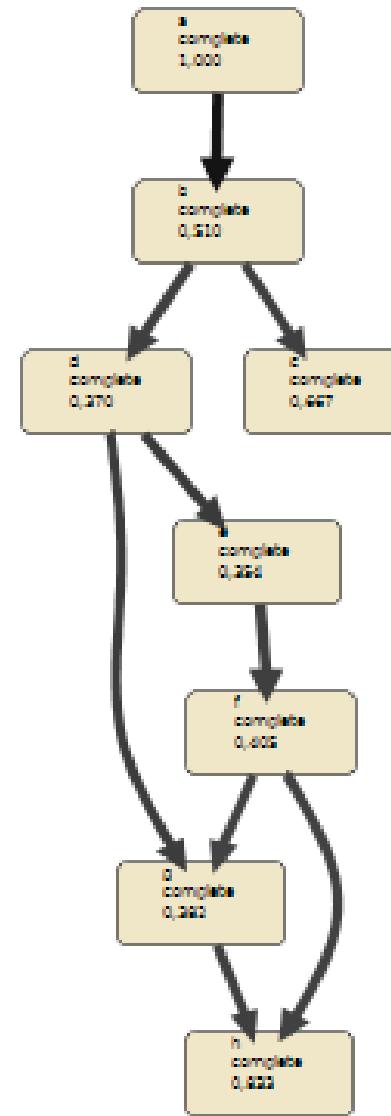
# Petri net - 3

- Исполнимы
- Очень похожи на конечные автоматы
- WF-сети: PN с одним началом и концом, связанные.
- Хорошие (sound):
  - Безопасные (safe): 1-bounded
  - Конец достижим
  - Завершимиые (кода отмечен конец, остальные места пусты)
  - Нет мертвых частей (можно прийти к активному состоянию перехода из любой достижимой разметки)

# Fuzzy net

- Места
- Дуги
- Без переходов
- Неисполнимы
- Удобны и наглядны

см. так же BPMN и примкнувший к ним UML



# Сопряженные процессы

- Больной в госпитале
  - Больной — кейс
  - Врач — место
  - Назначения и бумажки — токены
- Врач в госпитале
  - Врач — кейс
  - Больной — место
- Студенты и классы
- Разработчики и проекты
- Такие задачи мы сначала поручаем Васе, а если смог — Пете
- Таких разработчиков мы сначала отправляем на месяц в саппорты, потом в QA, а потом ставим на проекты.
- Ищем не правильную, а полезную точку зрения

# Логи



# Логи - 2

- Стартовая точка для майнинга процессов
- Все логи несчастливы по-своему
- Фильтрация и агрегирование
- Приведение к единому времени
- Есть ли начало и конец?
- Денормализованные логи
- Время — конец события?

# Логи - 3

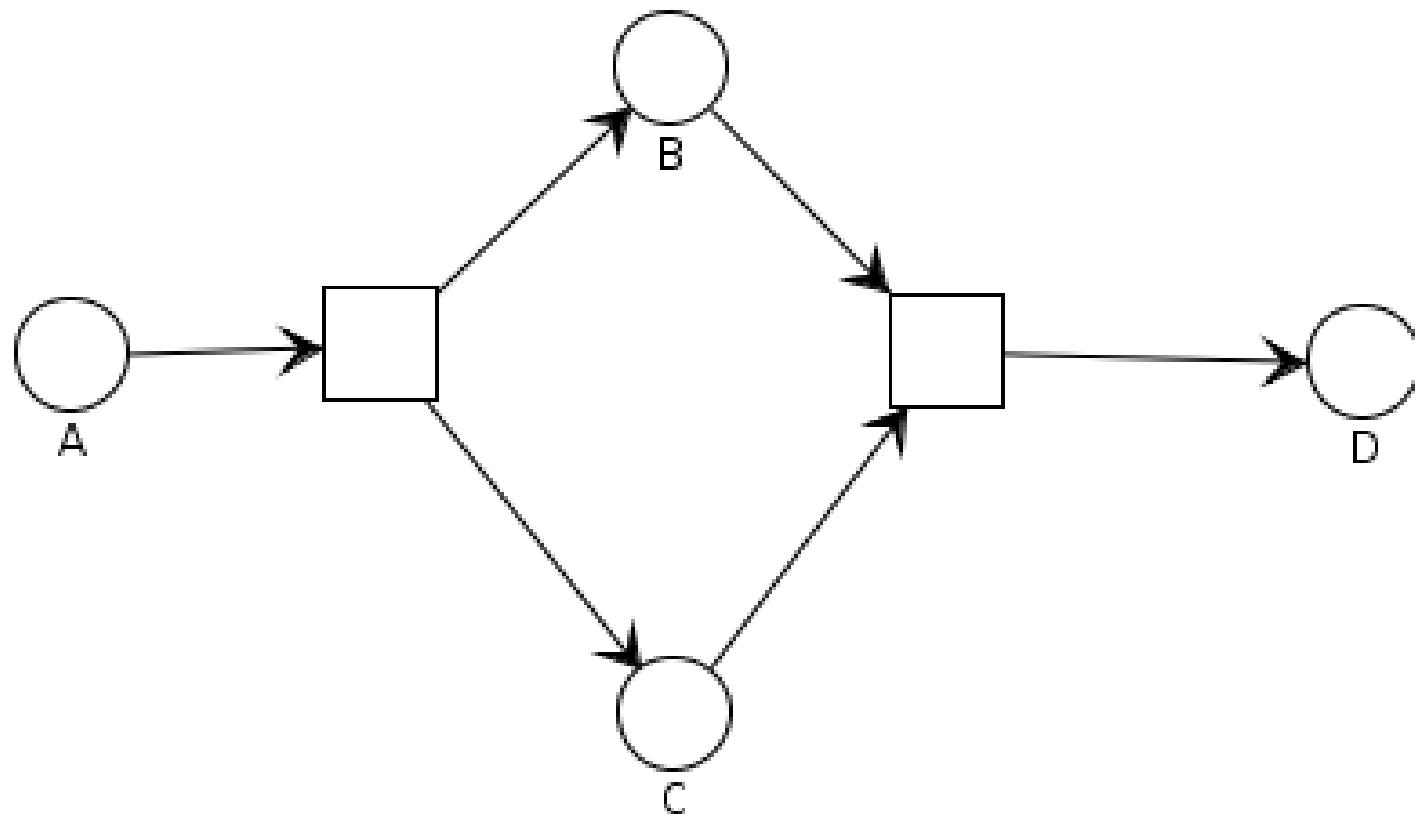
- Обычно доступно только событие завершения
- Время первого события не нулевое
- Иногда доступны время планирования, начала, конца.
  - конец в дело
  - остальное в подпроцесс
- Сколько ресурсов в событии?
- Иерархия процессов

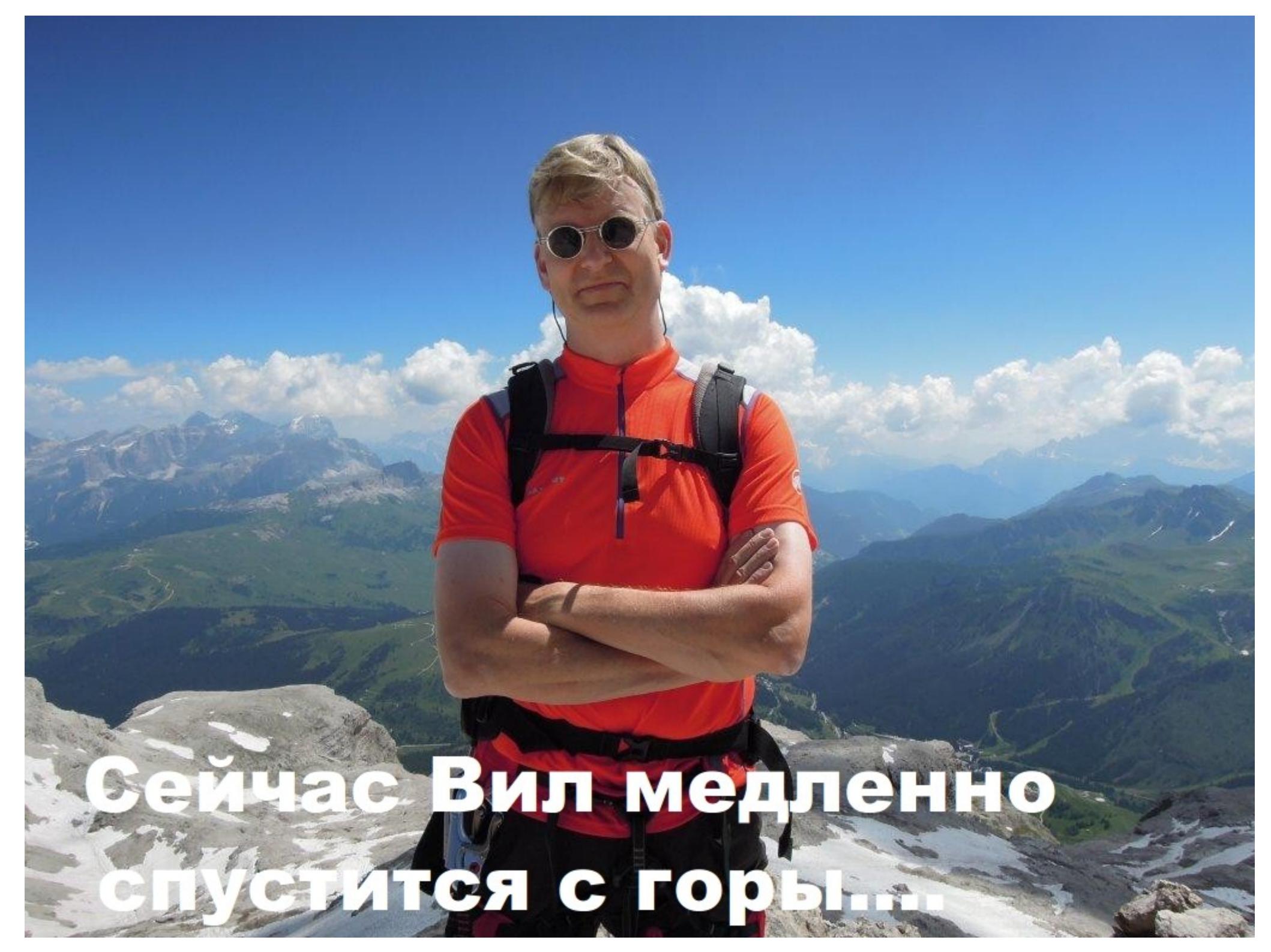
# Логи - 4

- IEEE 1849-2016 XES Standard
- IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams
- XML-based
- <http://pmuc.ing.ruc.cl/tools/opyenxes/>
- Хорошо, но опять где-то не у нас

# Параллельность

- $A \rightarrow B \rightarrow C \rightarrow D$
- $A \rightarrow C \rightarrow B \rightarrow D$





Сейчас Вил медленно  
спустится с горы....

Let  $L$  be an event log over  $T$ .  $\alpha(L)$  is defined as follows.

1.  $T_L = \{ t \in T \mid \exists_{\sigma \in L} t \in \sigma \},$
2.  $T_I = \{ t \in T \mid \exists_{\sigma \in L} t = \text{first}(\sigma) \},$
3.  $T_O = \{ t \in T \mid \exists_{\sigma \in L} t = \text{last}(\sigma) \},$
4.  $X_L = \{ (A, B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge \forall_{a \in A} \forall_{b \in B} a \rightarrow_L b \wedge \forall_{a_1, a_2 \in A} a_1 \#_L a_2 \wedge \forall_{b_1, b_2 \in B} b_1 \#_L b_2 \},$
5.  $Y_L = \{ (A, B) \in X_L \mid \forall_{(A', B') \in X_L} A \subseteq A' \wedge B \subseteq B' \Rightarrow (A, B) = (A', B') \},$
6.  $P_L = \{ p_{(A, B)} \mid (A, B) \in Y_L \} \cup \{ i_L, o_L \},$
7.  $F_L = \{ (a, p_{(A, B)}) \mid (A, B) \in Y_L \wedge a \in A \} \cup \{ (p_{(A, B)}, b) \mid (A, B) \in Y_L \wedge b \in B \} \cup \{ (i_L, t) \mid t \in T_I \} \cup \{ (t, o_L) \mid t \in T_O \},$  and
8.  $\alpha(L) = (P_L, T_L, F_L).$

# Ну он все равно не работает

- TL — все, кто встретился
- TI — стартовые
- TO — конечные
- $A > B$  (прямое следование)
- $A \rightarrow B$  iif  $A > B \wedge A \not< B$
- $A \# B$  iif  $A \not> B \wedge A \not< B$
- $A \parallel B$  iif  $A > B \wedge B < A$

# Ну он все равно не работает - 2

- Находим максимальные группы, такие что внутри группы они не следуют друг за другом, а между группами строго следуют
- Добавляем начало и конец
- Каждый элемент в (A, B) — место
- Соединяем дугами A — место — B
- Соединяем начало и конец
- Выдыхаем

# А если?

- Альфа\*
- Альфа+
- Альфа++ (серьезно)
- IPL
- Inductive Miner
- IPL Miner
- Тысячи их

# ProM и его друзья

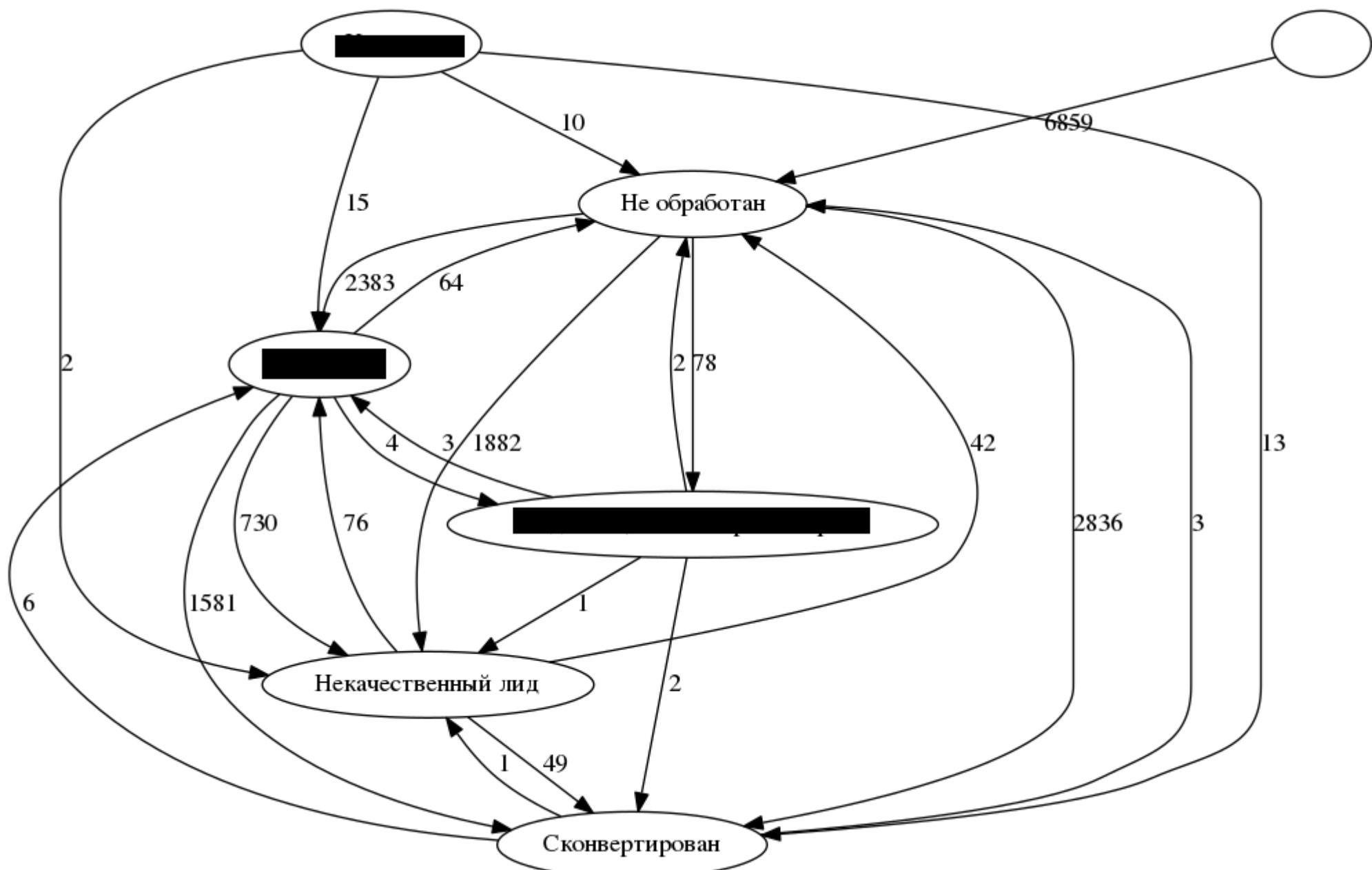
- ProM, без апдейтов
- Плагин для Rapid Miner
- Disco
- Все BI делают это, но как-то так
- ProM Lite

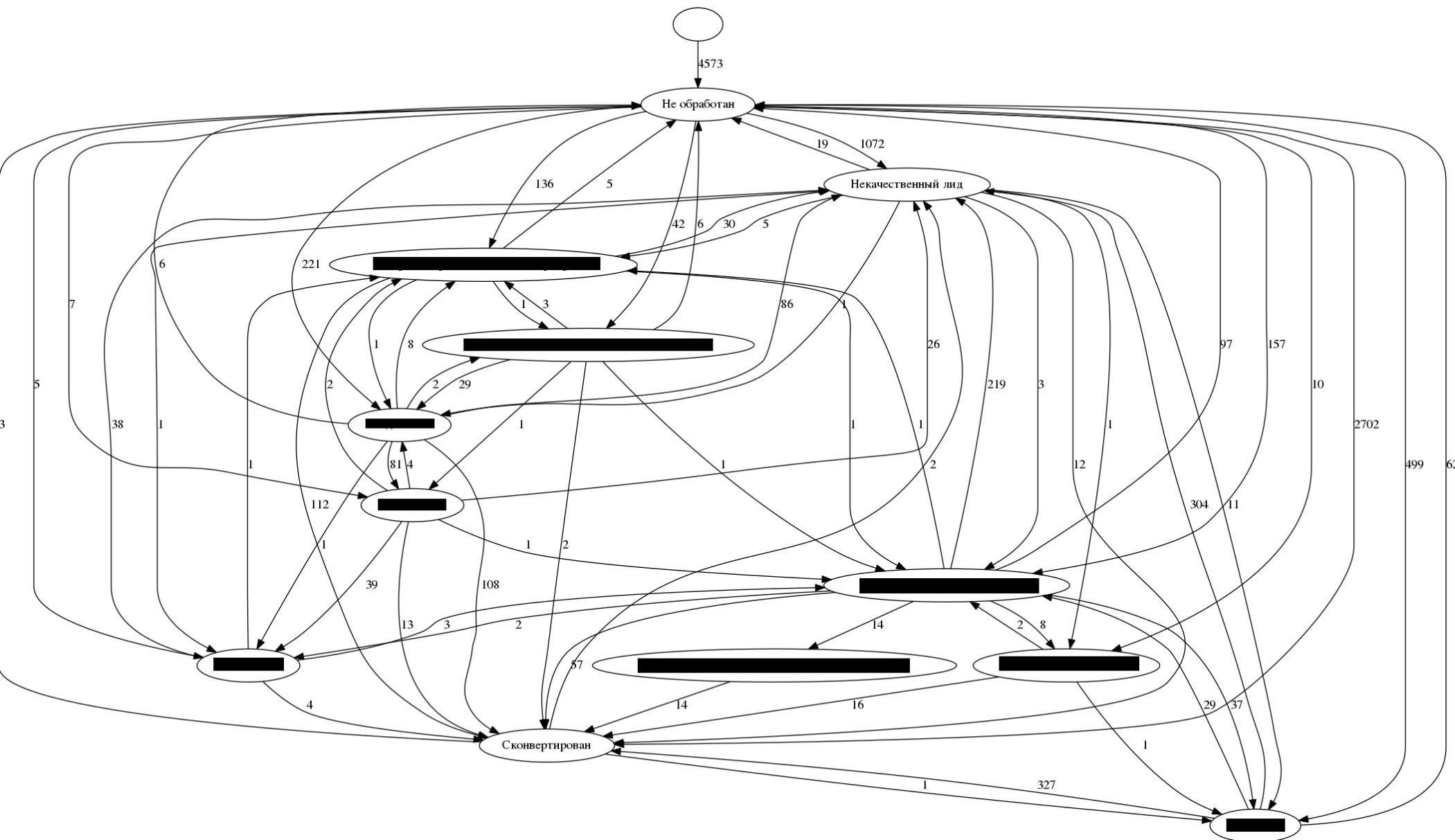
# ProM - 2

- Импортируем
- Конвертируем в XES
- XES - Альфа майнер - Петри нет
- Фильтрация логов
- Майнер эпизодов
- Fuzzy model
- Petri net, Inductive Miner, живой лог
- Dotted chart

# Что может быть интересно

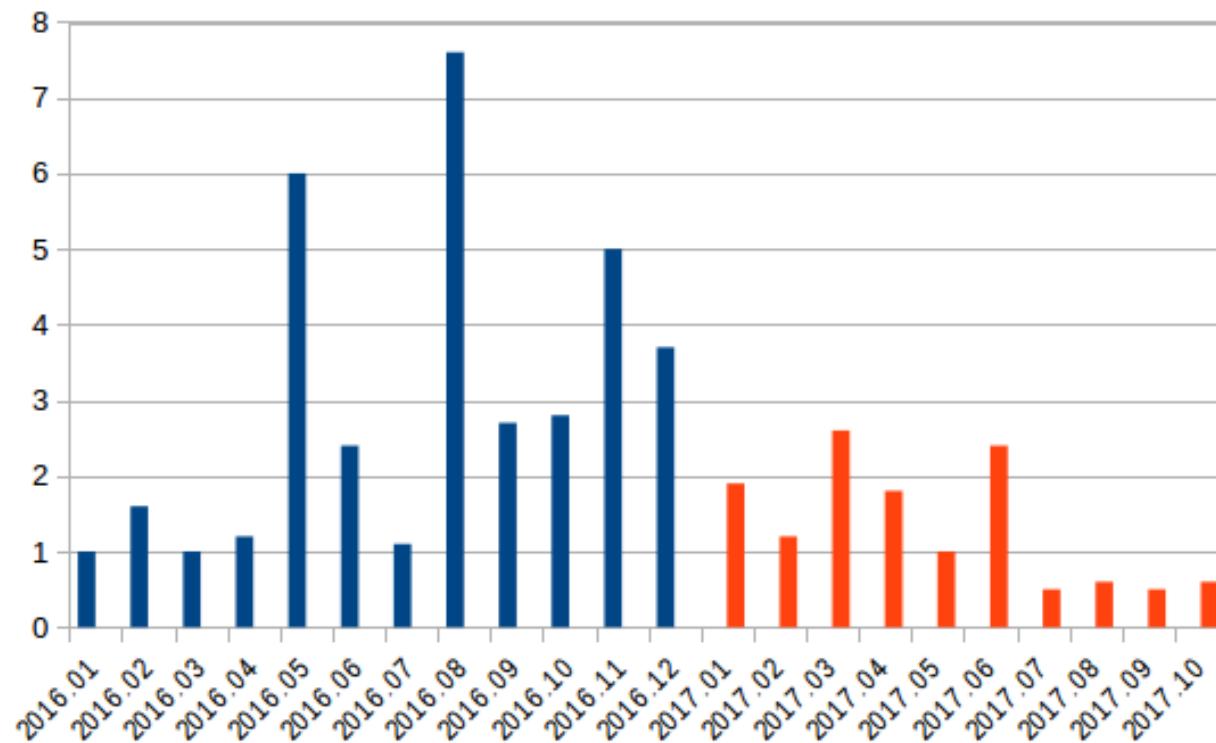
- Модель процесса
- Граф сотрудничества
- Распределение работы по сотрудникам
- Время перехода, гистограммы
- Очередь
- Диаграмма точками





# Найди 10 различий

- Изменения в процессе
- Ускорение обслуживания



# Что почитать

- Лекция Аалста в Яндексе
- <http://www.processmining.org/prom/start>
- Process Mining: Data science in Action
- Diogo R. Ferreira. A Primer on Process Mining  
Practical Skills with Python and Graphviz
- Wil M.P. van der Aalst. Process Mining
- Introduction to Event Log Mining with R

# За пределами WFMS логов

- Временные ряды, события серверных ферм.
- Логи высокой кардинальности (веб-сайты)
  - иерархия
  - группировка событий.
- События в жизни микрорайона
- Моделирование человеческой нервной системы как сети Петри

# Процесс как текст

- Сеть Петри как строка. См Visual Miner
- Sec2Sec преобразование
- Генерация обучающего набора

# Вопросы?

[dk.promsoft@mail.ru](mailto:dk.promsoft@mail.ru)

[@d\\_key](https://t.me/d_key)

<https://t.me/Promsoft>