

Фролов Ф.А.,
студент магистратуры,
2 курс, Институт экономики, математики
и информационных технологий,
факультет информационных технологий и анализа данных,
отделение Бизнес-информатики,
Российская академия народного хозяйства и государственной службы
при Президенте Российской Федерации,
Россия, г. Москва

Научный руководитель: Захарченко Ю.Ф., кандидат военных наук, доцент,
Доцент Кафедры системного анализа и информатики,
Российская академия народного хозяйства и государственной службы
при Президенте Российской Федерации,
Россия, г. Москва

**ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ УПРАВЛЕНИЯ БИЗНЕС-
ПРОЦЕССАМИ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ
PROCESS MINING**

***Аннотация:** В статье рассматривается технология из области управления бизнес-процессами, предназначенная для обнаружения, анализа и усовершенствования процессов бизнес-процессов в информационных системах на основании изучения системных данных из журналов событий.*

***Ключевые слова:** анализ процессов, большие данные, интеллектуальный анализ данных, бизнес-аналитика.*

***Annotation:** The article touches upon the issue of the technology in the field of process management that supports disclosing, analysis and improvement of business processes in information systems. The technology is based on event logs.*

***Key words:** process mining, big data, data mining, business intelligence.*

В начале XXI века человечество вступило в эпоху новой научно-технической информационной революции, это также период индустриальной революции четвертого уровня сложности, который заключается в применении современных информационных технологий для целей оптимизации производства товаров и услуг. Отчасти необходимость и важность оптимизации обусловлена глобализацией, которая способствует росту конкуренции. С другой стороны, информатизация общества и процессов способствует накоплению большого числа данных. Это было предсказано Гордоном Муром, основателем компании Intel и автором знаменитого закона Мура, согласно закону, мощность вычислительных устройств за относительно короткий промежуток времени может расти экспоненциально [1]. Мы наблюдаем экспоненциальный рост объема данных на протяжении последних 40 лет. На сегодняшний день задача состоит не в том, чтобы генерировать больше данных, а в том, чтобы превратить эти данные в реальную ценность. Еще одна особенность современной эпохи - скорость. Помимо того, что каждую секунду в мире генерируются большие объемы данных, сами данные непрерывно добавляются. Мы также наблюдаем разнообразие данных. Существует различные источники информации, данные представлены в различных форматах в цифровом виде. Не менее важно также учитывать и правдивость больших данных.

Таким образом, вопросы анализа данных возникают достаточно часто и являются актуальной задачей. Главная цель науки о данных (Big Data Science) – превращение данных в ценность [2].

Наука о данных позволяет найти ответы на вопросы о событиях, которые произошли в прошлом, проанализировать что именно произошло и найти причины возникновения событий. Помимо анализа прошлого, наука о данных также направлена ответить на вопросы о будущем: предсказать будущее и, если существуют такие возможности, улучшить или оптимизировать будущее.

Анализ бизнес-процессов включает в себя аналогичные задачи: предсказать, спланировать, оптимизировать или снизить издержки. Применение методов интеллектуального анализа к накопленным данным позволяет глубже

изучить и оптимизировать процессы. Процессно-ориентированный подход к науке о больших данных (Process Mining) позволяет сконцентрироваться не только на сборе и анализе данных, но и на анализе самих процессов, при этом Process Mining ориентирован на поиск новых решений и исправление проблем, отталкиваясь от реальных данных согласно установленным требованиям (Рисунок 1). В классическом интеллектуальном анализе данных люди обычно не занимаются анализом процессов. В областях же, где люди занимаются анализом моделей процессов, они обычно игнорируют данные.

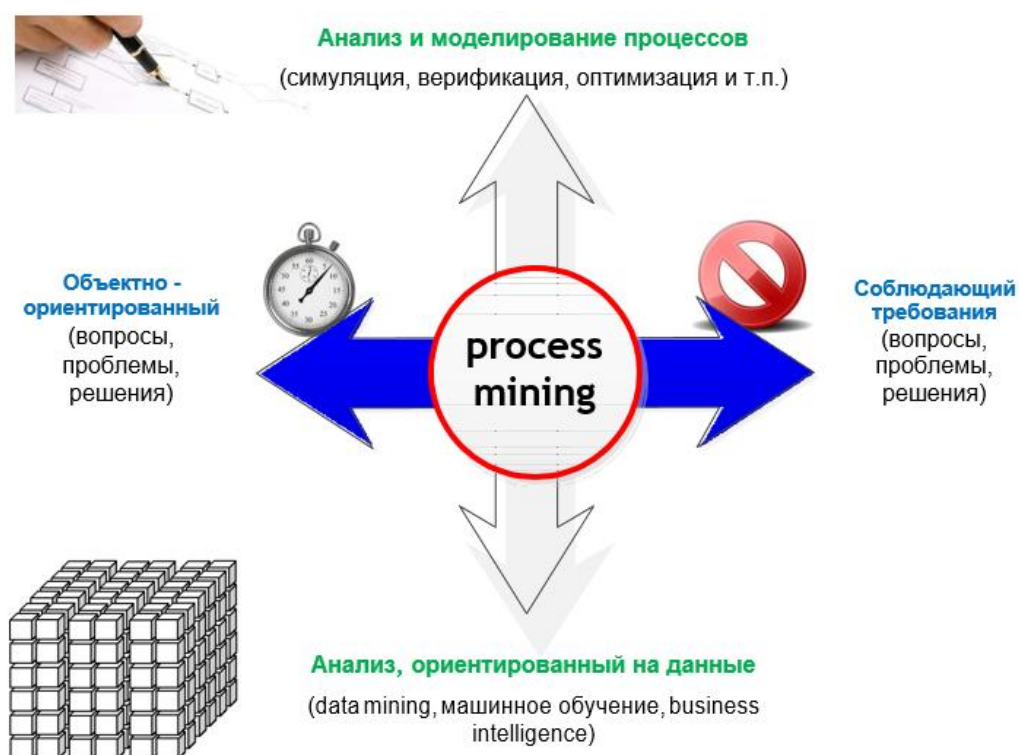


Рисунок 1. Сущность Process Mining

Таким образом, процессно-ориентированный подход к науке о больших данных преодолевает разрыв между классическим анализом данных и процессным моделированием.

Процессно-ориентированный подход к науке о больших данных фокусируется на процессах, используя реальные данные. Данные наблюдений фиксируются в журналах событий, каждая запись в журнале событий имеет определенный набор полей, описывающих это событие (Рисунок 2).

Обязательные поля журнала событий для Process Mining:

- case ID – id трассировки – набора событий, относящихся к отдельному конкретному экземпляру процесса;
- activity name – название активности, то есть отдельного события;
- timestamp – время фиксации события в журнале событий. Может указываться также два значения: время начала события и время окончания события, благодаря чему удастся вычислить длительность события и задержки между событиями.

Дополнительны поля в журнале событий для осуществления Process Mining могут включать в себя различную справочную информацию, например, на Рисунок 2 resource – имя сотрудника, который отвечал за осуществление события (activity name).

Журнал событий: обработка заказов					
номер заказа	активность	время	пользователь	продукт	кол-во
9901	зарегистрирован заказ	22-1-2016 09.15	Ольга Иванова	iPhone5S	1
9902	зарегистрирован заказ	22-1-2016 09.18	Ольга Иванова	iPhone5S	2
9903	зарегистрирован заказ	22-1-2016 09.27	Ольга Иванова	iPhone4S	1
9901	проверка запасов	22-1-2016 09.49	Петр Дуров	iPhone5S	1
9901	отправка заказа	22-1-2016 10.11	Ирина Ким	iPhone5S	1
9903	проверка запасов	22-1-2016 10.34	Петр Дуров	iPhone4S	1
9901	обработка оплаты	22-1-2016 10.41	Лилия Сорока	iPhone5S	1
9902	проверка запасов	22-1-2016 10.57	Петр Дуров	iPhone5S	2
9902	отмена заказа	22-1-2016 11.08	Лилия Сорока	iPhone5S	2

case id
 activity name
 timestamp
 resource
 other data

Рисунок 2. Журнал событий CRM системы

Существуют три типа отношений между моделями и данными о событиях: Play-out, Play-in, and Replay [3].

Play-Out – работа с готовой моделью, отталкиваясь от модели происходит моделирование поведения, то есть возможного сценария событий, противоположный тип отношений между моделью и данными о событиях – Play-In – построение модели базируется на фактических данных о событиях. На основе журнала событий, благодаря интеллектуальному анализу процессов, удастся обнаружить эти процессы и составить модель.

Тип отношений между моделью и событиями Replay означает возможность воспроизвести реальность поверх модели. Replay одновременно использует модель процесса (возможно, полученную при помощи Play-In) и данные в журнале событий (возможно, полученные при помощи Play-Out) для воспроизведения реальных последовательностей событий согласно модели. Благодаря вышеуказанным типам отношений между моделью и событиями, за счет применения Process Mining обеспечивается повышение эффективности управления процессами, а именно:

- удастся отследить отклонения в процессах;
- проанализировать производительность моделей, на основании фактических данных журнала событий появляется возможность отследить время, которое было потрачено на различные части процесса;
- на основании фактических данных определить временные задержки между отдельными составляющими процесса (Рисунок 3);
- благодаря многочисленным наблюдениям за трассировками, можно вычислить самые часто выполняемые события в процессе и отследить наиболее часто встречающиеся наборы событий в процессе;
- вывести из накопленных наблюдений распределение вероятностей;
- проанализировать самые короткие и быстрые пути между двумя событиями, если между ними существуют пути.

Ниже на Рисунок 3 представлена модель процесса с возможным набором событий, после отстройки модели и воспроизведения реальности поверх модели удалось вычислить продолжительность событий - числовое значение на синем фоне для каждого события, на черном фоне – временные задержки между событиями, тем самым удалось оценить временные затраты на выполнение процесса с указанным набором событий.

Replay с временными метками

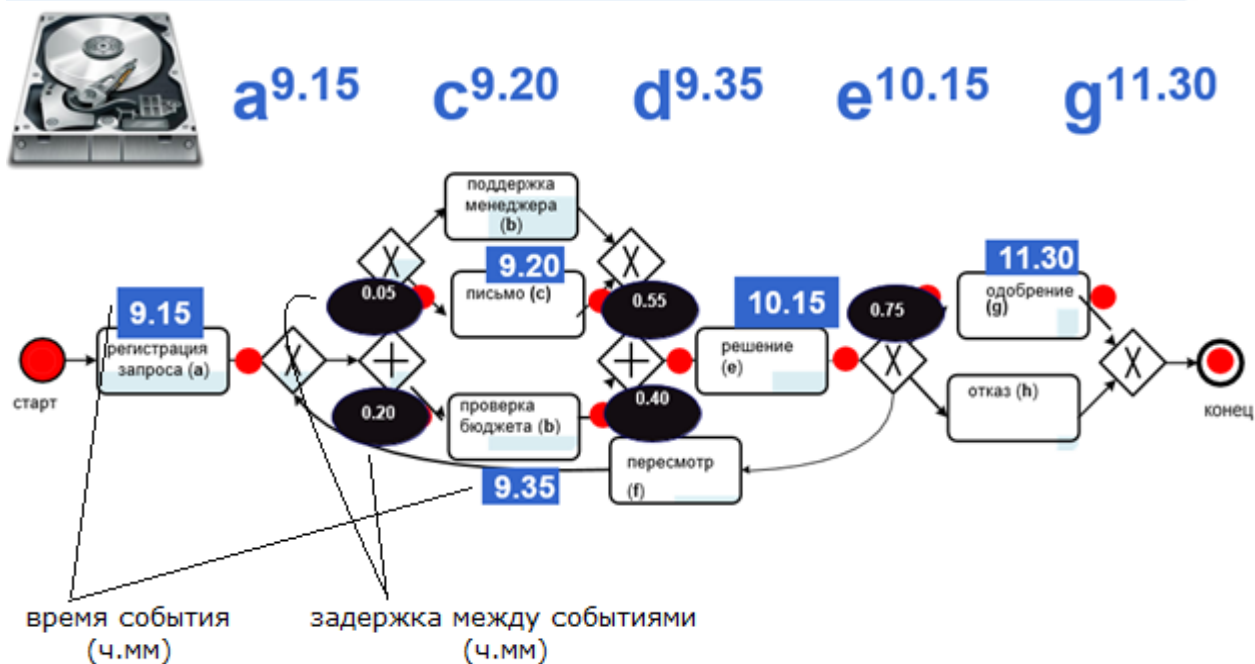


Рисунок 3. Возможный вариант течения процесса с временными характеристиками: время выполнения активностей и задержки между активностями

Процессно-ориентированный подход к науке о больших данных может быть полезен тем компаниям, в которых идут массовые процессы продажи и обслуживания, у которых есть большие обороты и большое количество контрагентов, в которых организованы длинные процессы, затрагивающие большое число функциональных подразделений. В 2017 году российский системный интегратор "Рамакс Интернейшнл" внедрил технологию Process Mining в банке "ВТБ 24". Решение о внедрении было принято с целью повысить эффективность работы розничной сети банка: увеличить производительность процессов, уменьшить трудозатраты на обслуживание клиентов-физлиц и сократить потери от продаж [4]. Условием для организации Process Mining является наличие программного обеспечения, которое бы фиксировало и накапливало происходящие события, данные могут быть взяты из различных источников, таких как: базы данных, журналы транзакций, ERP-системы, CSV-файлы, электронные таблицы Excel.

Если проводить сравнение с BI средствами анализа данных, как правило, благодаря таким средствам, как SAP BI или Oracle BI доступны к просмотру электронные таблицы, метрики, графики, то есть BI системы как бы пытаются “захватить реальность” в наборе чисел и расчетных ключевых показателей эффективности (KPI), чего порой бывает недостаточно. Этот факт наглядно иллюстрирует так называемый Квартет Энскомба: 4 различных диаграммы, включающие 4 различных набора данных, статистические свойства наборов данных (среднее значение переменной x, y ; дисперсия переменной x, y и т.д.) идентичны (Рисунок 4).

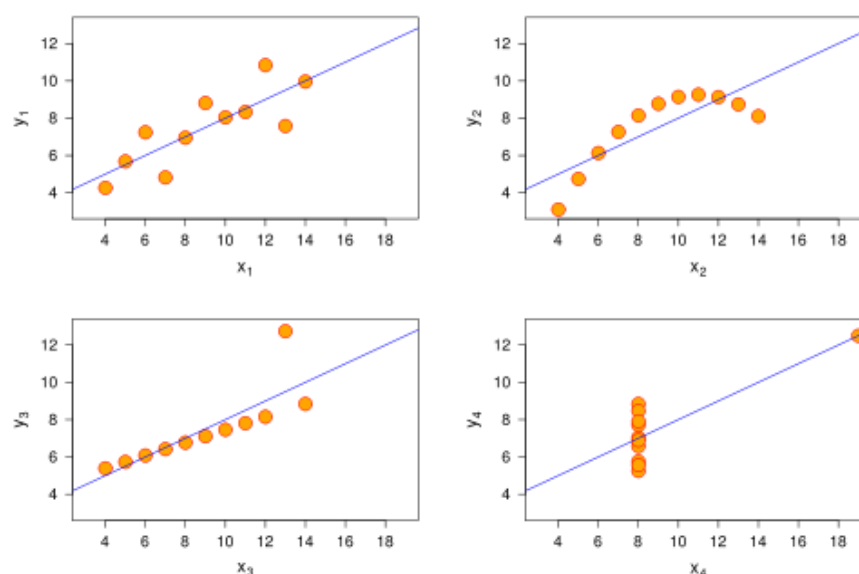


Рисунок 4. Четыре набора данных имеют идентичные статистические характеристики, но их графики существенно различаются.

Интеллектуальный анализ данных (Data Mining) работает с некоторыми типичными наборами данных, которые содержат в себе:

- числовые переменные, например, возраст или вес;
- категориальные переменные, которые не имеют числового значения, например, ‘правда’ или ‘ложь’. Категориальные переменные могут быть либо номинальные, либо порядковые.

Применительно к подобного рода данным существуют два типа методов интеллектуального анализа данных (Data Mining) с целью обнаружения внутренних взаимосвязей, закономерностей, существующих между объектами:

спонтанное машинное обучение (unsupervised learning) и принудительное машинное обучение с учителем (supervised learning). Для целей supervised learning поиска взаимосвязи между результирующей переменной и предикторной переменной используются такие алгоритмы, как: дерево принятия решений, регрессионный анализ.

Для целей unsupervised learning, где отсутствует понятие результирующей переменной, используются алгоритмы кластеризации и обнаружения шаблонов (pattern discovery). Для целей интеллектуального анализа данных используется специализированное программное обеспечение: RapidMiner, Weka, R, KNIME, SAS, QlikView, HANA, SAP BO. Существенное отличие Data Mining от Process Mining заключается в том, что все вышеперечисленные методы интеллектуального анализа данных и программное обеспечение, реализующее эти методы, не ориентированы на процесс. Такие задачи как обнаружение процессов, проверка соответствия, анализ узких мест не могут быть выполнены с использованием традиционных методов интеллектуального анализа данных.

Также стоит отметить, что разработанные инструменты Process Mining доступны для рядовых бизнес-пользователей. Обычно бизнес-пользователи имеют доступ к готовым дашбордам с ограниченным набором показателей, созданных под их профессиональные нужды. Excel также является достаточно популярным и мощным инструментом анализа и представления данных, но проанализировать процессы в Microsoft Excel проблематично, так как процессы состоят из большого числа событий. Стоит также отметить Microsoft Project, но данная программа заточена под управление проектами, которые состоят из набора последовательных или параллельных задач, включенных в проект и ведущих к конкретной цели, что существенно отличается от процесса [5]. Инструмент Process Mining Disco, разработанный в 2009 году, ориентирован на бизнес пользователей, которые, не зная никаких подробностей о базовых алгоритмах, могут применять его в работе с целью оптимизации рабочих процессов, главным условием является наличие журнала событий.

Для осуществления Process Mining недостаточно тех данных, которые применяются в Data Mining, так как там отсутствуют данные о событиях, Process Mining включает временной аспект и рассматривает выполнение одного процесса как последовательность выполненных действий. Но эти два подхода: интеллектуальный анализ процессов и интеллектуальный анализ данных могут быть объединены для того, чтобы ответить на сложные вопросы.

Стандарт XES (Extensible Event Stream) является стандартом, принятым целевой группой IEEE по интеллектуальному анализу процессов [6]. XES поддерживает общую структуру журнала событий: журнал (соответствующий процессу) содержит набор трассировок (т.е. конкретные экземпляры выполнения), каждый из которых, в свою очередь, содержит последовательность событий. В профессиональной среде существуют различные нотации для обозначения моделей процессов (UML, BPMN, сети Петри и т.п.), таким образом визуализировать модели можно различными способами для целей исследования и для представления конечных результатов.

На сегодняшний день для целей Process Mining широко используются следующие инструменты:

- ProM включает широкий функционал, который позволяет работать с различными типами моделей, а также позволяет использовать различные форматы входных данных [7].
- Disco является коммерческим инструментом, он прост, быстр и удобен в использовании и в основном использует нечеткие модели [8].
- Celonis – первое законченное приложение Process Mining корпоративного уровня, поставщиком данного программного обеспечения также является компания SAP. SAP Process Mining by Celonis существует в десктопной версии и в облачной версии [9].

Базовым алгоритм, заложенный в основу Process Mining, благодаря которому удается получить знания о структуре и поведении процесса из журналов событий, имеет название Альфа-алгоритм и является одним из первых алгоритмов, способных обнаруживать параллелизм и циклы [10]. Отправной

точкой для Альфа-алгоритма являются отношения упорядочения. Журнал событий может быть представлен в следующем виде:

$$L = [< abcd >^4, < acbd >^2, < aed >]$$

Где степени 2, 4 – количество одинаковых последовательностей.

На основании фактических данных определяются связи между событиями:

a, b, c, d, e. Связи могут быть четырех типов:

- прямой преемственности (обозначение “>”),
- причинности (обозначение “→”),
- параллелизм (обозначение “||”),
- отношения выбора (обозначение “#”).

На основании этих взаимосвязей модель процесса может быть сгенерирована автоматически по типу сети Петри (Рисунок 5).

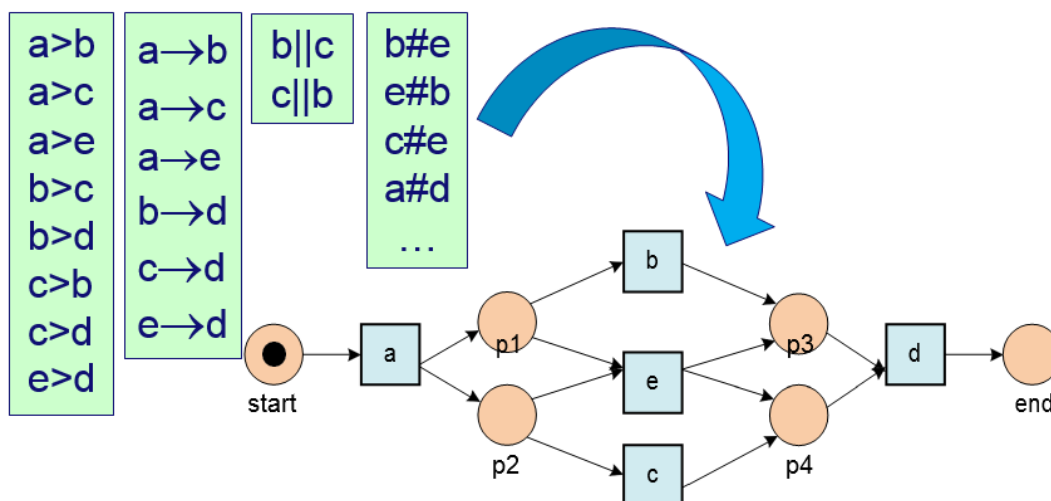


Рисунок 5. Сеть Петри, полученная в результате работы Альфа-алгоритма

Альфа-алгоритм состоит из семи шагов:

1. На первом шаге на основании журнала событий определяется конечный набор событий, которые будут присутствовать в сети Петри:

$$T_L = \{t \in T \mid \exists \sigma \in L \ t \in \sigma\}$$

2. Определяются начальные события на основании журнала событий, учитываются все последовательности:

$$T_I = \{t \in T \mid \exists \sigma \in L \ t = first(\sigma)\}$$

3. Определяются конечные события, те события, которые происходят в конце набора последовательностей:

$$T_0 = \{t \in T \mid \exists \sigma \in L \ t = \text{last}(\sigma)\}$$

4. Находятся множества событий A и B. События внутри каждого из множеств независимы друг от друга. Но между любыми двумя событиями из множеств A и B есть связь прямой преобладания, то есть в журнале наблюдений должна быть хотя бы одна позиция, в которой за элементом из множества A следует элемент из множества событий B, и это правило должно выполняться для всех комбинаций (Рисунок 6).

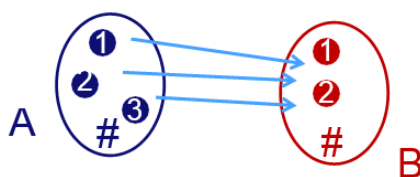


Рисунок 6. 4й шаг Альфа-алгоритма

5. Набор элементов во множествах A и B пересматривается, из него исключаются элементы, связи прямой преобладания которых повторно ссылаются на один и тот же элемент (Рисунок 7).

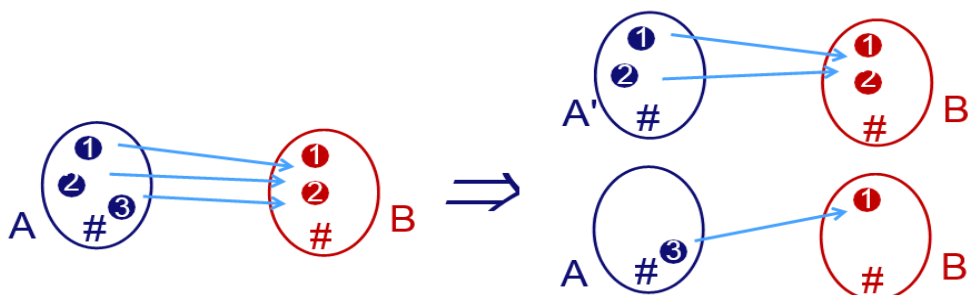


Рисунок 7. 4й шаг Альфа-алгоритма

6. К множествам событий из шага 5 добавляются: начальная, конечная и промежуточная позиция, на рис.5 они отмечены как: start, end, p1-p4.

7. На следующем шаге начинается отстройка переходов между элементами множеств A и B, начальной и конечной позицией.

У Альфа-алгоритма есть некоторые ограничения, например, в случае с указанными последовательностями событий модель будет строиться ошибочно:

- $L = [\langle abd \rangle, \langle abbbbd \rangle]$

- $L = [< abd >, < abc bcb d >]$.

Следующее ограничение на применение технологии Process Mining - это качество информации в журналах событий. Например, в электронной почте, где большинство сотрудников ведет деловую переписку, тем самым иницируя те или иные процессы и события к ним, проблематично собрать все письма в единый журнал наблюдения для построения модели, поскольку ни в теле письма, ни в его заголовке часто нет указания на то, к какому процессу и какому экземпляру соответствует данное письмо.

Именно поэтому технология Process Mining применима лишь там, где используются реляционные базы данных или создаются качественные журналы событий с указанием всех необходимых параметров. Еще одним ограничением является «правдивость» данных в информационных системах. На практике, после восстановления модели бизнес-процесса часто оказывается, что в информационной системе под чужим именем работал другой сотрудник, и само действие состоялось вчера, при этом информацию о нем внесли в систему только сегодня [11].

Таким образом, Process Mining является наиболее востребован там, где необходимо оптимизировать, проанализировать, ускорить массовые процессы продажи, обслуживания, сложные длинные процессы либо процессы, которые затрагивают большое число функциональных подразделений. Process Mining является современной цифровой технологией анализа процессов и помогает исследовать предметную область с достоверностью, на различных уровнях детализации, выявляя принципиально новые способы оптимизации процессов, а также незначительные отклонения, что необходимо для эффективного контроля, реагирования, принятия управленческих решения и, как следствие, создания конкурентного преимущества в современном мире.

ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Статья “Закон Мура” [Электронный ресурс]. URL: <http://cs.usu.edu.ru/study/moore/> (дата обращения: 07.01.2019)

2. Статья “Цифровая трансформация. Данные – ценность и актив компании” [Электронный ресурс]. URL: <https://itweek.ru/infrastructure/article/detail.php?ID=198324> (дата обращения: 07.01.2019)
3. Статья “Process Mining. Контролировать бизнес-процессы — легко!” [Электронный ресурс]. URL: <http://datareview.info/article/obzor-knig-process-mining-discovery-conformance-and-enhancement-of-business-processes/> (дата обращения: 07.01.2019)
4. Статья “Process Mining: "глаз" для обеспечения порядка” [Электронный ресурс]. URL: <https://www.comnews.ru/digital-economy/content/109715> (дата обращения: 07.01.2019)
5. Статья “Основное отличие процесса от проекта” [Электронный ресурс]. URL: rzbpm.ru/knowledge/upravlenie-processami-ili-upravlenie-proektami.html (дата обращения: 07.01.2019)
6. Официальный сайт стандарта XES. [Электронный ресурс]. URL: <http://xes-standard.org/> (дата обращения: 07.01.2019)
7. Официальный сайт поставщика ПО ProM. [Электронный ресурс]. URL: <http://www.processmining.org/prom/start> (дата обращения: 07.01.2019)
8. Официальный сайт поставщика ПО Disco. [Электронный ресурс]. URL: <http://fluxicon.com/disco/> (дата обращения: 07.01.2019)
9. Официальный сайт поставщика ПО Celonis. [Электронный ресурс]. URL: <https://www.celonis.com/solutions/> (дата обращения: 07.01.2019)
10. Описание Альфа-алгоритма. [Электронный ресурс]. URL: https://en.wikipedia.org/wiki/Alpha_algorithm (дата обращения: 07.01.2019)
11. Статья “Process mining – как увидеть фактические бизнес-процессы” [Электронный ресурс]. URL: <http://koptelov.info/publikatsii/process-mining-biznes-protsessy/> (дата обращения: 07.01.2019)