

Платунов Антон Игоревич

*студент бакалавриата, образовательная программа «Бизнес-информатика»,
Национальный исследовательский университет «Высшая школа экономики»,
г. Пермь*

E-mail: AIPlatunov@edu.hse.ru

Лядова Людмила Николаевна

*доцент кафедры математического обеспечения вычислительных систем,
Пермский государственный национальный исследовательский университет,
г. Пермь*

E-mail: LNLyadova@gmail.com

РАЗРАБОТКА КОНСТРУКТОРА ЖУРНАЛОВ СОБЫТИЙ С ДОПОЛНИТЕЛЬНЫМИ АТТРИБУТАМИ

Platunov Anton Igorevich

Student of the bachelor educational program “Business Informatics”

Higher School of Economics, Perm City

E-mail: AIPlatunov@edu.hse.ru

Lyadova Lyudmila Nickolaevna

Associate Professor of the Department of Computer Systems Software

Perm State University, Perm City

E-mail: LNLyadova@gmail.com

DEVELOPMENT OF A CONSTRUCTOR OF EVENT LOGS INCLUDING ADDITIONAL ATTRIBUTES

Аннотация: Цель проекта – разработка средств генерации и обработки журналов событий с использованием дополнительных атрибутов для анализа процессов с помощью средств Process Mining. Реализация основана на принципах low-code. Это обеспечивает пользователей, не владеющих навыками программирования, возможностью разрабатывать собственные правила обработки данных для формирования и предобработки журналов событий с дополнительными атрибутами. Ядро системы – многоаспектная онтология, включающая описания правил (функций) и данных. В исследовательском прототипе приложения реализованы алгоритмы разбора описаний правил и их интерпретации. Сформированные журналы сохраняются в форматах, используемых средствами Process Mining. Разработанный прототип иллюстрирует практическую значимость предложенного подхода.

Abstract: The goal of the project is to develop tools for generating and processing event logs including additional attributes for analyzing processes with Process Mining tools. Implementation are based on low-code principles. This enables non-programmers to develop their own data processing rules for generating and preprocessing event logs with additional attributes. The core of the system is a multifaceted ontology, including descriptions of rules (functions) and data. The research prototype of the application implements algorithms for parsing descriptions of the rules and their interpretation. Generated logs are saved in the formats used by Process Mining tools. The developed prototype illustrates the practical significance of the proposed approach.

Ключевые слова: журналы событий, атрибуты событий, анализ процессов, многоаспектная онтология, генерация журнала событий, конструктор правил, интерпретатор правил.

Keywords: event logs, event attributes, process mining, multifaceted ontology, event log generation, rules designer, rules interpreter.

Введение

Средства Process Mining используются в различных областях для анализа и оптимизации процессов на основе данных, которые, как правило, извлекаются из источников структурированных данных (журналов событий операционных систем и т. п.). В последние годы расширились возможности применения этих методов, в частности, для анализа процессов в медицине и пр. Помимо обязательных атрибутов событий, таких как идентификатор экземпляра (конкретной трассы), тип события (определяет конкретное действие, задачу) и метка времени, в источниках данных, используемых для анализа процессов, могут присутствовать дополнительные атрибуты (например, результаты лабораторных исследований, показатели, характеризующие состояние здоровья, и т. п.). Анализ динамики значений этих атрибутов позволяет получить важную информацию о процессах, сравнить эффективность тех или иных решений. Однако анализ процессов, выявление закономерностей с учётом значений дополнительных атрибутов требуют реализации новых методов, разработать которые можно с использованием языков программирования (например, плагинов на Java для платформы ProM). Разработка требует от исследователей, аналитиков (экспертов в соответствующей области) навыков программирования; или же к работе должны быть привлечены профессиональные программисты, способные создать приложения, реализующие новые методы анализа.

Цель проекта – апробация альтернативного подхода к созданию средств анализа процессов на основе журналов событий с использованием данных (дополнительных атрибутов), получаемых из различных источников, основанного на принципах low-code. Приложение должно включать конструкторы, с помощью которых пользователи, не владеющие навыками программирования, могли бы разрабатывать правила генерации журналов событий на основе данных различных типов, получаемых из выбранных ими источников, а также правила обработки

журналов событий с дополнительными атрибутами. Кроме того, должен быть реализован интерпретатор алгоритмов анализа, разработанных пользователями. Результаты обработки данных должны сохраняться в форматах, позволяющих реализовать взаимодействие разработанных средств с внешними системами, средствами анализа процессов (например, ProM).

Для достижения цели должны быть решены следующие основные задачи:

- 1) разработка способа представления правил (описаний функций);
- 2) разработка алгоритмов интерпретации правил (функций);
- 3) реализация пользовательского интерфейса приложения, доступного для использования специалистами, не владеющими навыками программирования.

Анализ процессов на основе гетерогенных данных: обзор литературы

В статьях [6, 7] рассматриваются вопросы анализа процессов с помощью методов, учитывающих значения дополнительных параметров событий, позволяющих выявить закономерности в динамике изменений значений их параметров. Авторами предлагаются определения событий, дополненных значениями атрибутов, операций агрегации и селекции атрибутов событий и пр. Предлагаемый подход к анализу реализован в приложении на языке Python.

Авторы [2, 3] также аргументируют необходимость разработки методов, которые нацелены не только на выявление закономерностей в потоках управления исследуемых процессов, но и на анализ потоков данных, автоматизацию выявления моделей данных в дополнение к автоматизации построения формальных моделей процессов.

В статье [8] представлены результаты аналитического обзора проблем и перспектив, связанных с генерацией журналов событий. Эксперты показывают, что особенно серьезные проблемы существуют в области интеграции данных в журналы событий. Авторы утверждают, что интеллектуальный анализ процессов может выиграть от интеграции методов анализа процессов и традиционных методов анализа данных.

Во многих статьях рассматриваются вопросы генерации журналов событий на основе данных, извлекаемых из различных источников: из баз данных и сообщений, передаваемых в системах торговых сетей [4, 5], из социальных сетей [9, 13, 14] и новостных лент [1, 15, 16]. Сформированные журналы могут быть дополнены атрибутами, которые также могут быть извлечены из указанных источников данных (например, в Интернет доступны данные о событиях, связанных с пандемией COVID-19, и данные о заболеваемости, связываемые с теми же объектами (странами, городами), что и события). При этом могут возникнуть проблемы отсутствия или неопределённости значений атрибутов, которые могут быть интегрированы в журналы событий и использованы в ходе анализа. Эти вопросы рассматриваются в статье [12]. Авторы приводят классификацию неопределённости и описывают способ представления атрибутов с неопределёнными значениями, показывают, как определить границы значений таких атрибутов при формировании моделей. Проблемам генерации журналов событий и их предварительной обработки посвящены также статьи [10, 11].

Однако в представленных исследованиях не рассматриваются вопросы разработки средств, которые позволили бы исследователям проводить эксперименты, не привлекая программистов для реализации предлагаемых методов на языках Java, Python и т. п.

Структура системы генерации и анализа журналов событий

Для реализации функций генерации журналов событий в структуру системы включены модули, показанные на рис. 1. Ядром системы является многоаспектная онтология, описывающая как предметные области, в которых проводятся исследования, так и используемые источники и структуры данных, а также функции, разработанные для генерации журналов и их обработки [17].

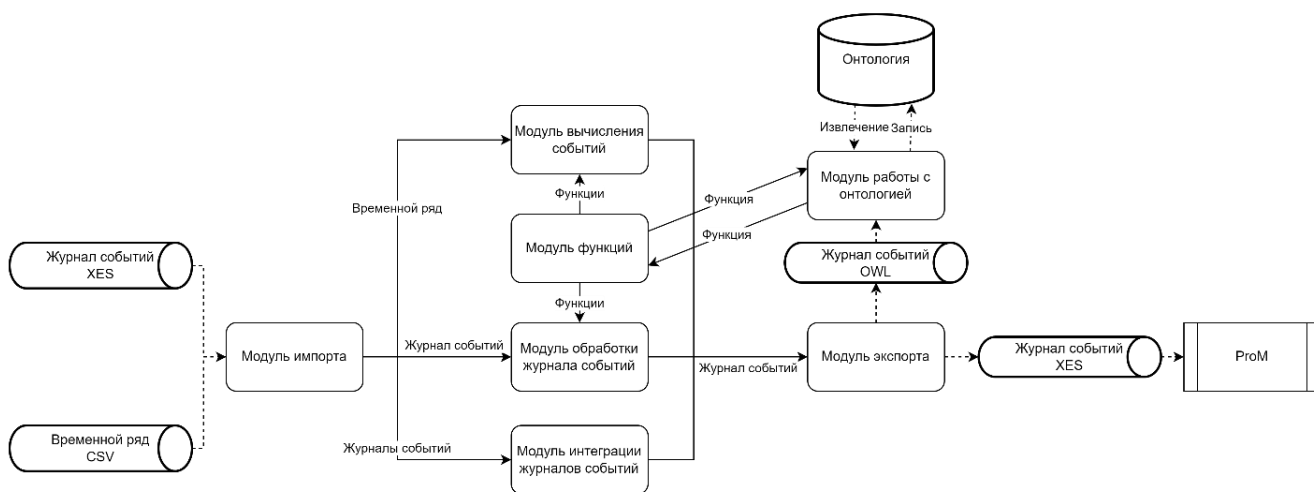


Рис. 1. Обобщённая структура системы генерации и анализа журналов событий

Сформированные журналы событий, которые могут содержать дополнительные атрибуты в соответствии с протоколом XES, также описываются в онтологии [15]. Через выгрузку сформированных журналов в формате XES возможно взаимодействие с системой ProM.

Представление правил генерации и обработки журналов

Требования к представлению правил (функций) генерации и анализа журналов событий определяются требованиями к пользовательскому интерфейсу, который должен обеспечивать простоту, доступность разрабатываемых средств для пользователей-непрограммистов: функции разрабатываются по единым правилам в конструкторах, имеющих унифицированный интерфейс.

В определении правил могут использоваться встроенные операции и функции, реализованные в системе и описанные в онтологии, операндами которых могут являться массивы значений, временные ряды, журналы событий и т. д., структура которых определена в онтологии, а также результаты вычисления операций и функций, т. е. правила – это суперпозиция функций.

В число встроенных (стандартных) включены и функции, которые позволяют определить альтернативные варианты вычислений, выполнять вычисления для заданных множеств объектов (например, записей в журналах

событий) или для диапазонов значений указанных атрибутов, реализованные по аналогии с соответствующими функциями MS Excel. Эти функции упрощают разработку алгоритмов для пользователей-непрограммистов: разработчики не должны изучать универсальные языки программирования и их конструкции, пользовательский интерфейс конструктора можно унифицировать, включив в него стандартные элементы управления для выбора при разработке алгоритма нужных функций и определения их параметров (аналогично средствам, которые используются при разработке формул в ячейках таблиц в MS Excel, или в построителе выражений в MS Access).

Для определения правил разработан язык, грамматика которого включает более 50 синтаксических правил. В языке определены, наряду со стандартными типами для языков программирования, типы данных для описания событий с дополнительными атрибутами и журналов событий, которые могут обрабатываться или генерироваться с помощью функций, временные метки, временные ряды и т. д. Целевым символом грамматики является Функция. Примеры определения нетерминальных символов в расширенной форме Бэкуса-Наура (БНФ), на основе которых разработан пользовательский интерфейс конструкторов, приведены ниже:

```

<Функция> ::= <Пользовательская функция> |
              <Функция генерации событий> |
              <Функция обработки событий>

<Пользовательская функция> ::= <Имя функции>
                               ( [ <Список формальных параметров> ] ) : <Имя типа> ;
                               ( <Выражение> )

<Список формальных параметров> ::= <Формальный параметр>
                                   { ; <Формальный параметр> }

<Формальный параметр> ::= <Имя параметра> : <Имя типа>

<Вызов функции> ::= <Имя функции> ( [ <Список фактических параметров> ] )

<Список фактических параметров> ::= <фактический параметр>
                                   { ; <фактический параметр> }

<Фактический параметр> ::= <Выражение>

```

Полужирным шрифтом в описании выделены терминальные символы.

Здесь при определении пользовательской функции могут использоваться параметры стандартных для языков программирования типов: числа, строки, массивы (в определении массивов пользователь использует базовые типы языка) и т. п. При вычислении выражения используются встроенные операции над данными стандартных типов (арифметические операции, операции над строками и т. п.) и встроенные функции (**MAX**, **MIN**, **AVG**, **SUM** и пр.). Для упрощения интерфейса конструктора выражений логические операции (**AND**, **OR**, **XOR**) реализованы как функции по аналогии с функциями MS Excel – это упрощает учёт приоритетов операций при формировании выражений и разборе сформированного пользователем определения функции. Конструктор выражений допускает также использование функций **SWITCH** или **IFTHEN** (для определения альтернативных вариантов вычисления в зависимости от заданных условий),

FOREACH (для обработки данных в цикле по заданным множествам объектов).

Правила описания функций генерации и обработки событий определяются аналогично, но для результатов выполнения этих функций устанавливаются типы **Event** и **EventLog**, так как эти правила определяют алгоритмы вычисления событий на основе данных, получаемых из различных источников, формирования журналов событий через интеграцию событий из нескольких журналов, которые могут быть сгенерированы с системе или импортированы из внешних источников, файлов. В качестве параметров функций могут быть заданы журналы событий (тип **EventLog**), временные ряды (тип **TimeSeries**), элементами которых являются числовые значения с временными метками, которые могут рассматриваться либо как значения дополнительных атрибутов, включаемые в записи формируемых журналов событий, либо как данные, на основе которых вычисляются значения дополнительных атрибутов или же типы событий, правила определения которых на основе анализа изменений этих значений определяют пользователи в описании правил (функций). Для этих типов данных определены встроенные функции **GETVAL** (выбор значения в указанном временном ряду по временной метке), **PREV** и **NEXT** (выбор предыдущего или следующего элемента временного ряда, журнала событий). Для вычислений значений дополнительных атрибутов или типов событий пользователь с помощью конструктора может построить выражения с использованием встроенных операций и функций или функций, определённых пользователем.

Описание правил на разработанном языке формируется с помощью конструктора (см. ниже). Далее выполняется разбор описания и формируется представление функции в онтологии рис. 2. Грамматика относится к классу $LL(1)$, разбор реализован на основе алгоритма леворекурсивного спуска.

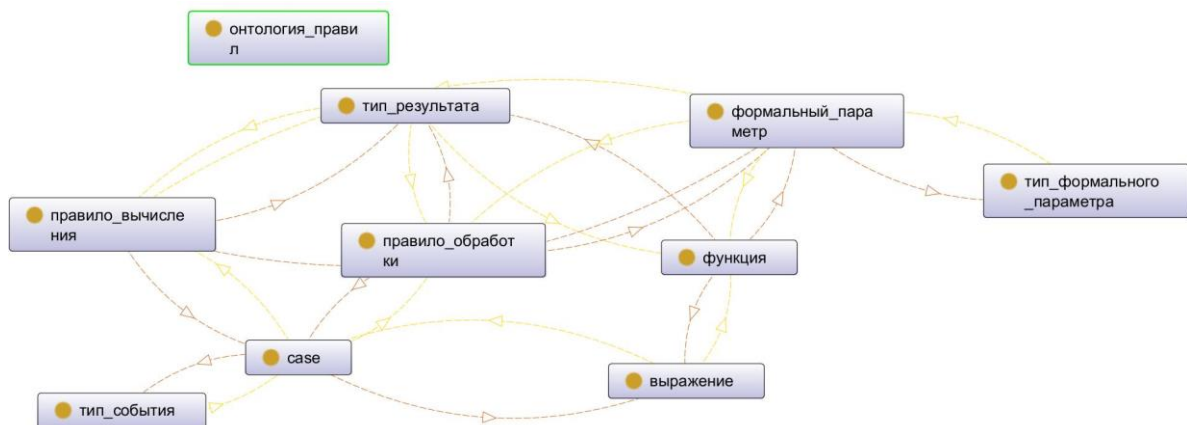


Рис. 2. Фрагмент онтографа, представляющий правило

На основе этих правил, определённых в грамматике, разработан унифицированный пользовательский интерфейс приложения (конструкторов), с помощью которого пользователи могут разработать свои функции, определяющие правила генерации и обработки журналов событий на основе данных, указанных в качестве параметров.

Алгоритм интерпретации функций (правил) реализуется на основе обхода онтографа, содержащего определение функции. Выбор функции и фактических

параметров (журналов событий, временных рядов, диапазонов значений и пр.) для её вычисления выполняется в окне интерпретатора правил.

Реализация конструктора и интерпретатора правил

На основе предложенного подхода разработан исследовательский прототип, реализующий основные функции системы:

– *Определение правил генерации событий и журналов событий*, в которые включаются дополнительные атрибуты, вычисленные по правилам, заданным пользователями (по определённым пользователям правилам формируются *событийные ряды* – журналы событий, расширенные дополнительными атрибутами, определение которых приведено в [9, 17]).

– *Интерпретация разработанных правил* – выполнение функций, определённых пользователями (формирование и предобработка журналов событий – подготовка журналов событий для анализа с помощью методов, реализованных в средствах Process Mining, например, в ProM и т. п.).

Форма конструктора правил показана на рис. 3. Пользователь может задать имя и описание для правила вычисления функции, результатом которой в данном случае является журнал событий, для которых определяются атрибуты по правилам, формируемым пользователем. Параметрами могут быть массивы значений, временные ряды, загружаемые из файлов и т. п. Способ вычисления определяется с помощью построителя выражений с использованием встроенных операций и функций, описания которых представлены в онтологии.

Идентификатор параметра	Тип параметра	Массив
timeseries_param	timeseries	<input type="checkbox"/>

Рис. 3. Форма конструктора правил (функций) генерации событий

Форма интерпретатора правил показана на рис. 4. Для апробации созданных средств разработаны правила формирования журналов событий на основе данных о заболеваемости COVID-19 в различных регионах России (для тестирования выбраны Москва и Санкт-Петербург) [17]. При определении параметров использована операция загрузки временного ряда из файла. На основе загруженных данных о заболеваемости в отдельном городе по правилу, заданному пользователем, сформирован журнал событий, в котором вычислен тип события. При загрузке данных по нескольким городам дополнительно можно было бы определить идентификатор, определяющий конкретную трассу по названию города. Далее можно было бы выполнить «слияние» журналов, сформированных по данным из различных источников с информацией по различным типам событий, например, проведённым в регионах мероприятиям, и выполнить анализ для выявления зависимости заболеваемости от этих мероприятий.

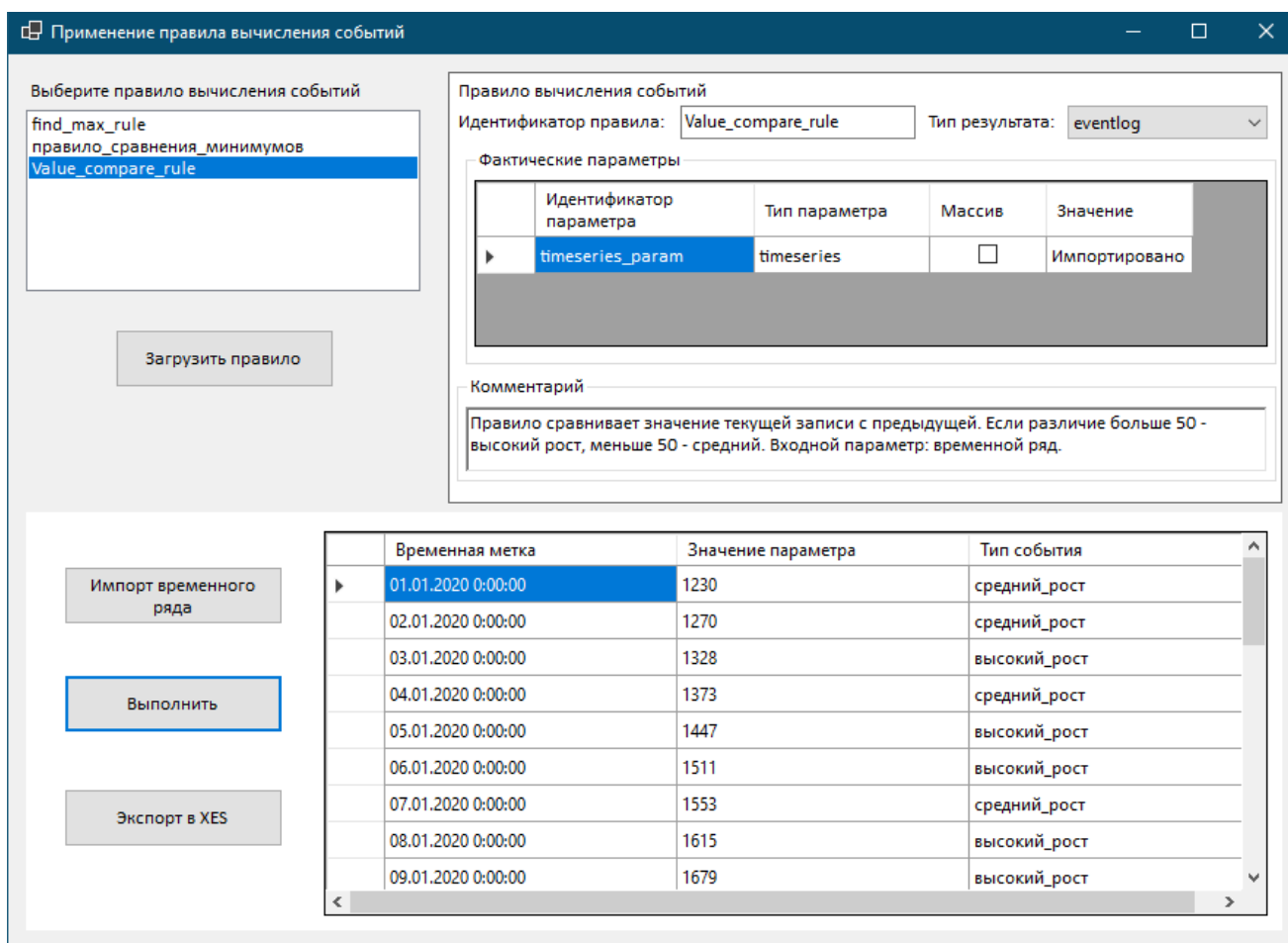


Рис. 4. Форма интерпретатора правил (функций) формирования журнала событий

Сгенерированный журнал может быть сохранён в файле в формате XES (рис. 5) и загружен, например, в систему ProM для анализа с использованием реализованных там методов. Возможность импорта и предобработки журналов, изменения их структуры и вычисления атрибутов по заданным правилам даёт возможность настроить журналы для применения различных средств Process Mining.


```

1 <?xml version="1.0" encoding="utf-8"?>
2 <log xes.version="1.0" xes.features="nested-attributes" xmlns:x="http://www.xes-standard.org/">
3   <extension name="Time" prefix="time" uri="http://www.xes-standard.org/time.xesext" />
4   <extension name="Concept" prefix="concept" uri="http://www.xes-standard.org/concept.xesext" />
5   <global scope="trace">
6     <string key="concept:name" value="__INVALID__" />
7   </global>
8   <global scope="event">
9     <string key="concept:name" value="__INVALID__" />
10    <string key="concept:eventType" value="__INVALID__" />
11  </global>
12  <string key="source" value="EventInterpetor" />
13  <trace>
14    <string key="concept:name" value="Case3.0" />
15    <event>
16      <string key="concept:name" value="UNDEFINED" />
17      <date key="time:timestamp" value="2020-01-01T00:00:00.000" />
18      <float key="float" value="1230" />
19      <string key="concept:eventType" value="средний_рост" />
20    </event>
21    <event>
22      <string key="concept:name" value="UNDEFINED" />
23      <date key="time:timestamp" value="2020-01-02T00:00:00.000" />
24      <float key="float" value="1270" />
25      <string key="concept:eventType" value="средний_рост" />
26    </event>

```

Рис. 5. Фрагмент кода журнала событий, сформированного в системе

Заключение

Подход к созданию гибких средств проектирования алгоритмов генерации и предобработки журналов событий с настройкой на потребности аналитиков, применение дополнительных атрибутов реализован в исследовательском прототипе. Он позволяет расширять набор доступных функций «на лету», проводить апробацию разработанных методов, вносить изменения в зависимости от результатов экспериментов. Применение не требует знания языков программирования – приложение доступно пользователям, знакомым со средствами MS Office.

Список литературы

1. Abrosimova P., Shalyaeva I., Lyadova L. The Ontology-Based Event Mining Tools for Monitoring Global Processes // Proc. of the IEEE 12th International Conference on Application of Information and Communication Technologies (AICT2018). – 2018. – P. 108-113. DOI: 10.1109/ICAICT.2018.8747094.
2. Bano D., Weske M. Discovering Data Models from Event Logs // Proc. International Conf. on Conceptual Modeling. ER 2020. Lecture Notes in Computer Science. – 2020. – V. 12400. – P. 62–76. DOI: 10.1007/978-3-030-62522-1_5.
3. Bano D., Zerbato F., Weber B., Weske M. Enhancing Discovered Process Models with Data Object Lifecycles // Proc. IEEE 25th International Enterprise Distributed Object Computing Conference (EDOC). – 2021. – P. 124–133. DOI: 10.1109/EDOC52215.2021.00023.
4. Calvanese D., Montali M., Syamsiyah A., van der Aalst W. M. P. Ontology-Driven Extraction of Event Logs from Relational Databases // Business Process Management Workshops. BPM 2016. Lecture Notes in Business Information Processing. – 2016. – V. 256. – P. 140–153. DOI: 10.1007/978-3-319-42887-1_12.
5. Carrasquel J. C., Chuburov S. A., Lomazova I. A. Pre-processing Network Messages of Trading Systems into Event Logs for Process Mining // Tools and Methods of Program Analysis. TMPA 2019. Communications in Computer and Information Science. – 2021. – V. 1288. – P. 88–100. DOI: 10.1007/978-3-030-71472-7_7.

6. Cremerius J., Weske M. Change Detection in Dynamic Event Attributes // Business Process Management Forum. BPM 2022. Lecture Notes in Business Information Processing. – 2022. – V. 458. – P. 157–172. DOI: 10.1007/978-3-031-16171-1_10.
7. Cremerius J., Weske M. Context-Aware Change Pattern Detection in Event Attributes of Recurring Activities // Intelligent Information Systems. CAiSE 2023. Lecture Notes in Business Information Processing. – 2023. – V. 477. DOI: 10.1007/978-3-031-34674-3_1.
8. Kampik T., Weske M. Event Log Generation: An Industry Perspective // Enterprise, Business-Process and Information Systems Modeling. BPMDS EMMSAD 2022. Lecture Notes in Business Information Processing. – 2022. – V. 450. – P. 123–136. DOI: 10.1007/978-3-031-07475-2_9.
9. Lanin V., Lyadova L., Zamyatina E., Vostroknutov N. An Ontology-Based Approach to Social Networks Mining // Proc. 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. – 2021. – V. 2: KEOD. – P. 234-239. DOI: 10.5220/0010716600003064.
10. Marin-Castro H. M., Tello-Leal E. Event Log Preprocessing for Process Mining: A Review // Applied Sciences. – 2021. – 11(22). – 10556. DOI: 10.3390/app112210556.
11. Mitsyuk A. A., Shugurov I. S., Kalenkova A. A., van der Aalst W. M. P. Generating event logs for high-level process models // Simulation Modelling Practice and Theory. – 2017. – V. 74. DOI: 10.1016/j.simpat.2017.01.003.
12. Pegoraro M., van der Aalst W.M.P. Mining Uncertain Event Data in Process Mining // Proc. International Conference on Process Mining (ICPM). – 2019. – P. 89–96. DOI: 10.1109/ICPM.2019.00023.
13. Peña-Araya V., Quezada M., Poblete B. Galean: Visualization of Geolocated News Events from Social Media // Proc. of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'15). – 2015. – P. 1041–1042. DOI: 10.1145/2766462.2767862.
14. Ritter A. Open domain event extraction from twitter // Proc. of the 18th ACM SIGKDD international conference on Knowledge Discovery and Data Mining. – 2012. – P. 1104-1112. DOI: 10.1145/2339530.2339704.
15. Shalyaeva I., Lyadova L., Lanin V. Events Analysis Based on Internet Information Retrieval and Process Mining Tools // Proc. of 10th International Conference on Application of Information and Communication Technologies (AICT2016). – 2016. – P. 168–172. DOI: 10.1109/ICAICT.2016.7991678.
16. Shalyaeva I., Lyadova L., Lanin V. Ontology-Driven System for Monitoring Global Processes on Basis of Internet News // Proc. of IEEE 11th International Conference on Application of Information and Communication Technologies (AICT2017). – 2017. – P. 385–389. DOI: 10.1109/ICAICT.2017.8687086.
17. Zayakin V., Lyadova L., Sminov M., Lanin V., Matta N., Zamyatina E. Event Series Generation and Analysis Based on Multifaceted Ontology // Proc. of the IEEE 16th International Conference on Application of Information and Communication Technologies (AICT2022). – 2022. – 6 pp. DOI:10.1109/AICT55583.2022.10013573.