The practical application of tools and process mining methods

A A Stupina\textsuperscript{1,2,3}, A V Kukartsev\textsuperscript{1,2}, O A Antamoshkin\textsuperscript{1,2,3}, E D Agafonov\textsuperscript{1,2}, N V Fedorova\textsuperscript{1,2} and E G Korepanova\textsuperscript{2}

\textsuperscript{1} Siberian Federal University, 79, Svobodny pr., Krasnoyarsk, 660041, Russian Federation
\textsuperscript{2} Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation
\textsuperscript{3} Krasnoyarsk State Agrarian University, 90, Mira Av., Krasnoyarsk, 660049, Russian Federation

E-mail: h677hm@gmail.com

Abstract. The article discusses the implementation of Process Mining methods in the activities of educational institutions. Process Mining should be used in those moments when it concerns the activities of participants in information systems. The sequence of stages is checked very carefully, and the productivity of each participant, division of the organization is monitored. There are doubts that it will not be possible to change the mechanics of the process, but it will be possible to save money on debugging the work of each link, which together will give a tangible effect. These are difficult, long-term, peculiar processes that involve many departments. In this case, Process Mining is no longer looking for a way to accelerate, but a way to simplify the process, rationalize the number of steps, avoid process malfunctions, unnecessary cycles, involving unnecessary participants, etc. Ideal and real models of the course were drawn up and compared.

1. Introduction

Today, due to the pandemic, e-courses are the main platforms for teaching university students. In order for distance learning students to fully acquire the knowledge that they would have received during a visit to the institute, teachers need to pay special attention to the composition of their course. Often, institutions conduct various surveys for students to assess the quality of the course. Often, it is better to analyze the real process of student interaction with the course.

The purpose of this work is to analyze the electronic course "Design and architecture of information systems" using the process mining technology.

2. Materials and Methods

Process analysis (process mining) is a method that allows you to extract a model of a business process based on information collected during the actual execution of the process [1]. The main advantage of process analysis is its objectivity - this method does not use an idealized model. Process mining tools discover actual process models based on raw data from event logs. In other words, the technology is able to collect information from various ISs used at the enterprise (1C, IP-telephony, HR systems), and build an accurate map of any process.

There are three approaches to process mining:
Extract. The starting point for the process analysis is to obtain an event log, which can be represented by a file with a sequence of events, an Excel table, a database table, etc. Each event is a line in the log that contains the process ID, event name, date and time of the occurrence of this event. Based on the event log, a model is created without using a priori information. At this stage, various algorithms are used. An example is the \( \alpha \)-algorithm. This algorithm takes the event log and creates a Petri net based on the entries in the event log [2].

Verification of compliance. At this stage, the existing process model is compared with the event log to verify that it is consistent with the reality obtained from the log. Thus, compliance checking can be used to detect, localize and explain the detected deviations [2].

Improvement. The goal of this step is to improve the existing process model using information about the actual process recorded in the event log. One of the types of improvement is changing the prior model that is, modifying the model to better reflect reality. Another type of improvement is expansion, that is, adding a new perspective to the process model by cross-correlating it with the journal [2].

At the moment, events are registered in many business processes, corporate systems, medical systems, automation and management systems, social networks and other areas, which allow the use of process mining in these areas for better understanding, analysis and improvement of processes.

The main industries that use process analysis are healthcare, information technology, industry, education, finance and logistics [3].

In healthcare, the use of process analysis can help determine patient flows for specific diseases, treatment chances, maximizing flow with good outcomes, correlations between prescribed treatments, quality of treatment, and complications. With the help of process analysis, hospital organizations can evaluate how clinical trials are conducted, check whether certain clinical guidelines and medical protocols are actually being followed, visualize resource use, and identify bottlenecks [4].

Process analysis is also a useful tool for examining clinical pathways describing the flow of patient attendance according to their illness. In addition, process analysis is used in emergency medicine - an area dedicated to the diagnosis and treatment of unforeseen diseases and injuries. Thus includes an initial patient assessment to check for complications, diagnosis, treatment and coordination of care between different medical departments suitable for a patient in need of surgical or non-surgical care [5].

In information technology, process analysis is used for such purposes as agile task management, collaboration, productivity improvement, and verification of compliance with operational and service management requirements. In addition, this method is used to increase the demand for an application or site among users by changing the design or functionality of the software [6].

Process analysis helps in the detection and monitoring of educational processes. Particular attention is paid to finding suitable learning paths based on the profile of a group of students, defining learning styles and the desired type of content, and finding learning paths to achieve the best results, analyzing student relationships [7].

Process analysis is also used in finance for risk analysis, insurance claims processing, automated ATM process analysis, contractual security review, fraud and root cause analysis, bank contact center efficiency improvements, loan approvals, and credit card or clerk checks [8].

Process analysis is used in logistics processes associated with improving warehouse layout, forecasting wholesale ports, detecting train rerouting, freight traffic anomalies, transportation planning and resource allocation in an aircraft manufacturer [9-11].

Also, process analysis is applied in the following areas:

- Social sphere: public administration, government, municipal services, social services and postal services.
- Consulting: bureau of consulting services related to consulting, audit and taxes.
- Biology: processes in biological systems such as metabolomics, biochemical networks, etc.
• Entertainment: Apps for sports, games, advertising and media.
• Retail: online and offline stores.
• Agriculture and food industry.

The application of process mining in the field of education is shown in Figure 1. This diagram is an adaptation of the general one to this area.

Figure 1. Application of process mining in education.

Field of Education: Students and teachers are an integral part of e-learning. The task of teachers is to provide various resources for teaching students (courses, lectures, tests, practical works, exams).

Learning environment: various devices (computers, laptops, tablets, phones, etc.) through which the learning process takes place directly. The environment not only provides the necessary resources, but also records the events that occur during the learning process [12, 13].

Event logs: files that store events that occur during training (files with a sequence of events, Excel tables, database tables, etc.).

Process Models: Models created from information collected from event logs and reflecting how e-learning users interact with the system [10].

3. Result
The starting point for conducting a process analysis is to obtain data from the event log. Electronic courses that are used at the Siberian Federal University use the Moodle system.

Moodle is a free learning management system focused primarily on organizing interaction between teacher and learners, although it is also suitable for organizing traditional distance courses, as well as supporting face-to-face learning [11].

The following information is recorded in the event log of the Moodle system (Table 1):

• Date and time of the event.
• IP address of the course participant's computer.
• Last name, first name and patronymic of the course participant.
• An action performed by a course participant.
• Additional information.
Table 1. Moodle system Event Log.

| Time            | Full username | Event context                  | Component | Event name | Description       | A source | IP address |
|-----------------|---------------|--------------------------------|-----------|------------|-------------------|----------|------------|
| 23/12/20, 19:09 | Student 1     | Assignment: Laboratory work No. 11 - DFD methodology | File response | Answer provided | The user with | web | 183.12, 4.231.119 |
| 23/12/20, 19:09 | Student 1     | Assignment: Laboratory work No. 10 - Using UPD categories | File response | Answer provided | The user with | web | 183.12, 4.231.119 |
| 22/12/20, 18:54 | Student 2     | Test: Practice Skill Test -2 | Test       | Test attempt started | The user with | web | 10.30, 105.38 |
| 22/12/20, 18:51 | Student 2     | Test: A test of practical knowledge - 8 | Test       | Test attempt started | The user with | web | 10.30, 105.38 |
| 22/12/20, 18:43 | Student 2     | Lecture: Lecture 8 (high trajectory) - IDEF3 requirements for describing business processes | Lecture | Lecture started | The user with | web | 10.30, 105.38 |
| 22/12/20, 13:40 | Student 3     | Task: Task LW 6 (high trajectory) | File response | Answer provided | The user with | web | 37.112, .195.1, 25 |
| 22/12/20, 13:40 | Student 3     | Task: Task LW 5 (high trajectory) | File response | Answer provided | The user with | web | 37.112, .195.1, 25 |
| 21/12/20, 16:38 | Student 2     | Lecture: Lecture 3 (high trajectory) - Quality management systems. Fundamentals and vocabulary | Lecture | Lecture started | The user with | web | 10.30, 105.38 |
| 21/12/20, 16:33 | Student 2     | Test: Practical Knowledge Test - 2 | Test       | Test attempt started | The user with | web | 10.30, 105.38 |
| 21/12/20, 16:02 | Student 2     | Assignment: Laboratory work No. 12 - Reengineering processes | File response | Answer provided | The user with | web | 10.30, 105.38 |
| 13/12/20, 23:44 | Student 3     | Test: Practice Skill Test - 2 | Test       | Test attempt started | The user with | web | 37.112, .195.1, 25 |

As a rule, the sequence of actions of students does not correspond to the ideal model of the course, drawn up by the teacher.

Using the information obtained from the event log, the teacher can create a model of real student behavior using various applications. For the analysis of the electronic course “Design and architecture of information systems” three groups were selected, for each of the groups a real model of the course by students was created using the “Celonissnap” system. An example of a model for one group of students is shown in Figures 2-5 (sheet 1-4) [12].
Figure 2. Standard real model for the first group of students, sheet 1.

Figure 3. Sheet 2.
Analyzing the ideal and real obtained models, it becomes obvious that the real process of completing the course is significantly different from the ideal. In this course, there are three trajectories for different levels of knowledge: low, medium, high, but group students use only a high trajectory, which means that perhaps the presence of a low and medium trajectory does not make sense or the entrance test that determines the trajectory for a student is not complex enough [14]. In addition, this course contains sections for loading practical work that do not belong to any of the trajectories, and students upload some of their work to these sections, which causes difficulties for the teacher when checking the work and determining which trajectory this student belongs to. Also, general lectures in this course are not viewed by most of the students. If the teacher considers these lectures important, then you should add test questions to them for compulsory passage, or close access to the assignments located after this lecture until it is viewed [11].
4. Conclusion

Thus, teachers can visually study the created model of student behavior, thereby observing not only the learning outcome, but the entire process. By analysing a real-life course model based on information obtained from event logs, an instructor can optimize their course by modifying some of the course elements [14].

References

[1] Repin V V 2013 Business processes. Modeling, implementation, management (Moscow: Mann, Ivanov and Ferber) p 512

[2] Kukartsev V V, Khramkov V V, Fedorova N V, Rozhkov A V, Tynchenko V S and Bashmur K A 2020 Features of evaluating the effectiveness of industrial enterprise marketing activities IOP Conference Series: Materials Science and Engineering 734(1) 012081

[3] Btemirova R I, Lazarova L B, Kairova F A, Sopoeva I A and Dov I G 2020 Project-based method in the organization of educational activities Journal of Physics: Conference Series 1691(1) 012191

[4] Milov A V, Tynchenko V S, Kukartsev V V, Tynchenko V V and Antamoshk in O A 2018 Classification of non-normative errors in measuring instruments based on data mining Advances in Engineering Research 158(2) 432-7

[5] Van der Aalst W 2011 Process mining: discovery, conformance and enhancement of business processes Springer Science & Business Media 55(8) 76-83

[6] Becker J, Delfmann P, Eggert M and Schwittay S 2012 Generalizability and Applicability of Model-Based Business Process Compliance-Checking Approaches – A State-of-the-Art Analysis and Research Roadmap BuR Business Research Journal 5(2) 221-47

[7] Trachuk A V, Linder N V, Tarasov I V, Nalbandyan G G, Khovalova T V, Kondratyuk T V and Popov N A 2018 Transformation of industry in the conditions of the fourth industrial revolution: monograph (Moscow: Real economy) p 147

[8] Golubitskaya N, Kosheleva T and Kunin V 2017 Problems of innovative development of an entrepreneurship in the industry in the conditions of upgrade of Economy IOP conference series: Earth and environmental science 90(1) 012049

[9] Kukartsev V V, Chzhan E A, Tynchenko V S, Antamoshin O A and Stupina A A 2018 Development of adaptive educational course in the SibFU E-learning system J. of Siber. Fed. Univer. Humanities and Social Sci 11(5) 740-52

[10] Tynchenko V S, Tynchenko V V, Bukhtoyarov V V, Kukartsev V V, Kukartsev V A and Ereemeev D V 2019 Application of Kohonen self-organizing maps to the analysis of enterprises’ employees certification results IOP Conf. Ser.: Materials Science and Engineering 537(4) 042010

[11] Nurrhaman A A, Husen N P and Rukmana O 2020 Designing Information System for Student Practicum Assessment in the Laboratory IOP Conference Series: Materials Science and Engineering 847(1) 012047

[12] Boyko A A, Kukartsev V V, Smolina E S, Tynchenko V S, Shamlitskiy Ya I and Fedorova N V 2019 Imitation-dynamic model of amortization of reproductive effect with different methods of calculation J. Phys.: Conf. Ser 1353(1) 012124

[13] Feoktistov A V, Trofimenko O N, Ognev S P, Lyakhovets M V and Koynov R S 2020 Digital educational platform as a personnel management tool Journal of Physics: Conference Series 1691(1) 012067

[14] Reimann P, Markauskaite L and Bannert M 2014 e-Research and learning theory: What do sequence and process mining methods contribute? British Journal of Educational Technology 45(3) 528-40