INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2024, VOL. 70, NO. 4, PP. 389-396
Manuscript received October 26, 2023; revised June, 2024. doi: 10.24425/ijet.2024.149557

# Formation of Models for Registering Systemic Processes in The Digital Educational Environment of the University Based on Log File Analysis

Valerii Lakhno, Bakhytzhan Akhmetov, Kaiyrbek Makulov, Bauyrzhan Tynymbayev, Svitlana Tsiutsiura, Mikola Tsiutsiura, and Vitalii Chubaievskyi

***Abstract*—It has been demonstrated that technologies and methods of intelligent data analysis (IDA) in the educational domain, particularly based on the analysis of digital traces (DT) of students, offer substantial opportunities for analyzing student activities. Notably, the DT of students are generated both during remote learning sessions and during blended learning modes. By applying IDA methods to DT, one can obtain information that is beneficial for both the educator in a specific discipline and for the educational institution's management. Such information might pertain to various aspects of the functioning of the digital educational environment (DEE) of the institution, such as: the student's learning style; individual preferences; the amount of time dedicated to a specific task, among others. An algorithm has been proposed for constructing a process model in the DEE based on log analysis within the DEE. This algorithm facilitates the description of a specific process in the DEE as a hierarchy of foundational process elements. Additionally, a model based on cluster analysis methods has been proposed, which may prove beneficial for analyzing the registration logs of systemic processes within the university's DEE. Such an analysis can potentially aid in detecting anomalous behavior of students and other individuals within the university's DEE. The algorithms proposed in this study enable research during log file analysis aimed at identifying breaches of information security within the university's DEE.**

***Keywords*—digital educational environment; digital footprints; log files; data mining (or intellectual data analysis); algorithm; model**

## I. Introduction

A natural consequence of the advancement of information technologies (IT) is their global integration into all existing aspects of human activity. The contemporary individual, to varying degrees, coexists in two realms – the tangible and the virtual. While the former has been the subject of study by scholars in the fields of social, humanities, natural, and to some extent, technical sciences throughout human civilization's existence, the latter space is still in the phase of recognizing the need for its exploration and gradual comprehension.

As it becomes an integral part of societal life, the virtual domain increasingly intertwines with our daily existence. Concurrently, the potential risks associated with IT amplify. This pertains, in particular, to processes of personal data storage and processing, corporate network security, and protection against cybercriminal intrusions. Consequently, for companies, enterprises, and governmental institutions (generally referred to as entities of informational activity or EIA), it is imperative to understand how to ensure a high level of information security (IS) and maintain control over their information assets (IA). These challenges have also permeated the educational sector. Specifically, a key challenge in education is the processing and analysis of vast data volumes generated by higher educational institutions. Information about students, curricula, educators, and learning processes continuously accumulates in educational management systems. Under such conditions, the information systems (IS) of universities become a source of the so-called "digital traces" (DT). In other words, data characterizing a user's network activity. Automated logging of actions by educators and students in the IS and the university's network unveils new research horizons. For instance, analyzing this data using artificial intelligence methods can identify patterns, trends, and critical factors influencing educational quality. However, there's a need for the development and application of innovative methods and tools for the efficient processing and analysis of such data volumes. All the aforementioned points underscore the relevance of new research in the domain of methods and models for analyzing the DT of university and college students.

Valerii Lakhno is with National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine (e-mail: lva964@nubip.edu.ua).
Bakhytzhan Akhmetov is with Abai Kazakh National Pedagogical University, Almaty, Kazakhstan (e-mail: bakhytzhan.akhmetov.54@mail.ru).
Kaiyrbek Makulov and Bauyrzhan Tynymbayev is with Caspian University of Technology and Engineering named after Sh.Yesenova, Almaty, Kazakhstan (e-mail: kaiyrbek.makulov@yu.edu.kz, b.tynymbayev@yu.edu.kz).
Svitlana Tsiutsiura, Mikola Tsiutsiura and Vitaliy Chubaievskyi are with State University of Trade and Economics, Kyiv, Ukraine (e-mail: stsutsura@knute.edu.ua, mtsiutsiura@knute.edu.ua, chubaievskyi_vi@knute.edu.ua).

## II. Literature review

In recent years, a considerable number of scientific publications have been dedicated to the methods and tools used for analyzing the digital traces (DT) of university students.

In study [1], the authors conducted research related to the development of tools for collecting data categorized as DT. The experience of using digital media was examined, encompassing over 500 students from Turkish universities. Based on the data processing from the experiment, the authors conclude that online tools leaving the maximum amount of digital traces include online chats, social networks, emails, web pages, and cloud document-sharing environments. A set of tools leaving the least amount of digital traces was also identified, which includes discussion forums and learning management systems. However, the study focuses on experimental setup and data processing without proposing new models and methods for analyzing student DT.

In [2], the authors emphasize that educational technology specialists should provide participants in the educational process with information regarding the assessment of active and passive DT. Such information dissemination, for instance to students, will, in the authors' opinion, minimize potential negative consequences from data leaks. The work is primarily informative and does not cover all aspects of studying DT in university and college information systems (IS).

In [3], the authors investigate DT issues in automated design systems popular among designers. The results of this study allow for a quantitative assessment of student performance during design project implementation based on DT analysis. However, the authors limited their scope to a narrow segment of application software, not addressing a broader range of software used during student education.

In [4], the authors explored the system architecture of a student relationship management system based on Internet of Things technologies for collecting DT in educational institutions. A system architecture for the proposed Student Relationship Management System (SRMS) was introduced. The results obtained allow educational institution management to receive analytical data to enhance the quality of educational services and facilitate the analysis of student behavior in higher educational institutions.

In [5], the authors noted that the digital transformation of society, including the educational sector, has introduced radical adjustments to existing learning models. According to the authors of [5], this includes models of electronic education and procedures for forming DT of students and educators.

As demonstrated in studies [1, 2], within the context of analyzing the educational process, digital traces (DT) can be interpreted as an electronic representation of data on the educational, professional, and social activities of students and educators. According to [1, 2, 3], DT in educational institutions characterizes the level of professional competencies of students and educators.

Typically, the following components of DT are distinguished:

1) 1. Technical-technological components: These reflect human activities in the digital space from the perspective of utilizing recording technologies such as log files, IP addresses, visited and requested web page addresses, access point identifiers, user-entered biometric and personal data, data and parameters of protocols used for information exchange (e.g., TCP/IP, HTTP, FTP, POP3, SMTP, TELNET), and others.
2) 2. Personal-psychological components: These reflect the social profiles of specific individuals, including personal social media accounts, online posts, photos, comments, etc.
3) 3. Behavioral components: These reflect actions undertaken by individuals, which might be initiated by personal perceptions of the world and/or their social attitudes. Examples include search engine queries, online orders of goods and services, geolocation system-recorded movements, etc.
4) 4. Activity components: These encompass data on digital media, such as presentations, projects, recorded videos, photos, audio data, etc.
5) 5. Competency components: These reflect the knowledge and skill levels of an individual. For educational institutions, this might include grades for academic subjects, reviews on completed coursework and thesis projects, feedback on abstracts, articles, etc., and electronic certificates and diplomas (e.g., from Coursera, Edx).
6) 6. Communicative components: These reflect the communication system within the educational and professional environment of an institution, such as participation in forums, chats, corporate email correspondence, etc.
7) 7. Reflective components: These represent self-reflection and the results of self-analysis of one's educational and professional activities. This pertains to self-assessment as an individual, one's actual and potential capabilities, etc. Examples include responses to various surveys and questionnaires, sociological research data, etc.

It's evident that the aforementioned components are closely interrelated. However, the degree of DT fixation varies for each component. While for some components, DT can be recorded and analyzed using already tested technologies, such as processing log files (technical-technological components), other components require the application of more resource-intensive technologies, like Data Mining (intelligent data analysis technologies).

Regarding the educational process in universities and colleges, collecting and analyzing digital traces (DT) of students and educators is not an end in itself. This procedure for collecting and analyzing DT aims to adjust the educational trajectory of development and improve the quality of education.

Many studies in recent years have been dedicated to the application of intelligent analysis technologies for processing educational data.

As shown in studies [6-16], the DT of university and college students can be considered a source of data on the level of student competencies, strategies for their further development, and subsequent professional activities.

In [6-10], it is noted that the DT of university and college students begin to form essentially at the stage of the admissions campaign and upon admission to the educational institution. Information about student activities, even in the minimal DT configuration, contains personal data, information about the specialty and chosen educational trajectory, academic performance, time spent in distance learning systems, and other details.

The authors of works [11-16, 23-24] note that technologies and methods of intelligent data analysis (IDA) in the field of education, including based on the analysis of DT of students, offer great opportunities for analyzing student activities. Moreover, DT of students are formed both during distance learning and during blended learning. By applying IDA methods to DT, useful information can be obtained for both the educator of a specific discipline and the management of the educational institution. Such information may concern, for example, aspects such as the student's learning style, personal preferences, the amount of time spent on a specific task, etc. Ultimately, this kind of information contributes to making a more accurate forecast of student performance and can provide reasoned recommendations to enhance the effectiveness of the educational process.

In summary, it can be stated that the development of models and methods of intelligent data analysis concerning the digital traces of students and educators in the digital educational environment (DEE) of educational institutions is a relevant task, and some aspects of its solution are the subject of this work.

## III. THE PURPOSE OF THE STUDY.

Is to develop methods and models for data analysis that are formed in the digital educational environment (DEE) of educational institutions based on digital traces. Such studies will help identify the relationships between the structure and content of educational programs (EP) and syllabi with the results of students' educational activities, ultimately allowing for the optimization of the educational process.

## IV. METHODS AND MODELS

As the analysis of studies dedicated to the application of Data Mining (DM) technologies in the educational process has shown, there remain issues primarily related to the insufficient elaboration of data interpretation methodologies obtained from different components of the DEE. This includes data related to the digital traces of students and teachers.

Let's consider the main information sources and data flows in the university's educational process that relate to the formation of digital traces, see Fig. 1.

Within the scope of this article, we will only consider the aspect touching upon the technical and technological components of digital traces in the DEE. Specifically, we address the task of constructing a model of processes in the DEE. Many processes occurring in the DEE require accounting for the basic elements of log files, which, for instance, capture the behavior of processes in the DEE (loading new materials by teachers, using virtual machines (VMs) for performing educational tasks, submitting test and control assignments, etc.).

Partially, the initial data for the intelligent data analysis system in the DEE can be event logs. These logs will then reflect the sequences of actions when implementing a specific process, for example, launching and using a VM by a student and/or teacher. Based on the merging and analysis of logs, it will be possible to construct a multivariate model of the VM launch process in the DEE.

The resulting hierarchical model [17] should consist of levels corresponding to the hierarchy of performers and reflecting the interrelation of the process actions based on its basic elements.

In the corresponding log files, tasks performed on the VM will be recorded, system status changes will be displayed, as well as changes in its individual components, etc.

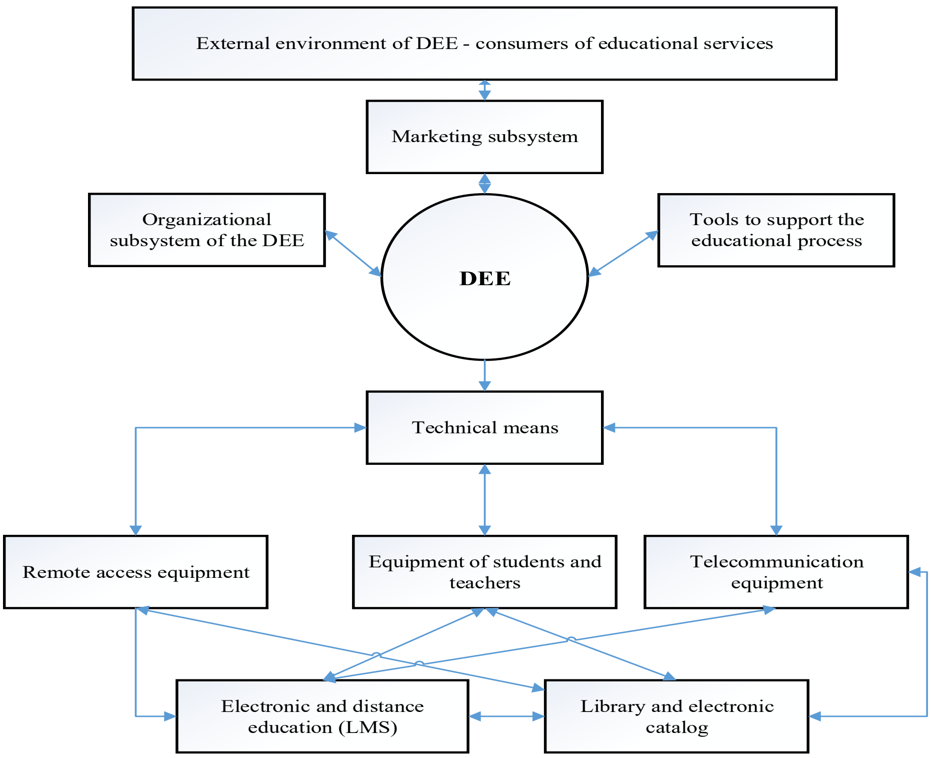


Fig. 1. Main information sources and data flows in the educational process

Within the scope of our research, an algorithm for constructing a model of processes in the Digital Educational Environment (DEE) was proposed. The foundation of the suggested algorithm is based on the tree of a specific process within the university's DEE. Utilizing such a hierarchical model provides DEE administrators with information concerning individual branches of processes. This, in turn, facilitates the elimination of various process variations, optimizing the data for subsequent intellectual analysis.

The algorithm is depicted in Figure 2 in a general form.

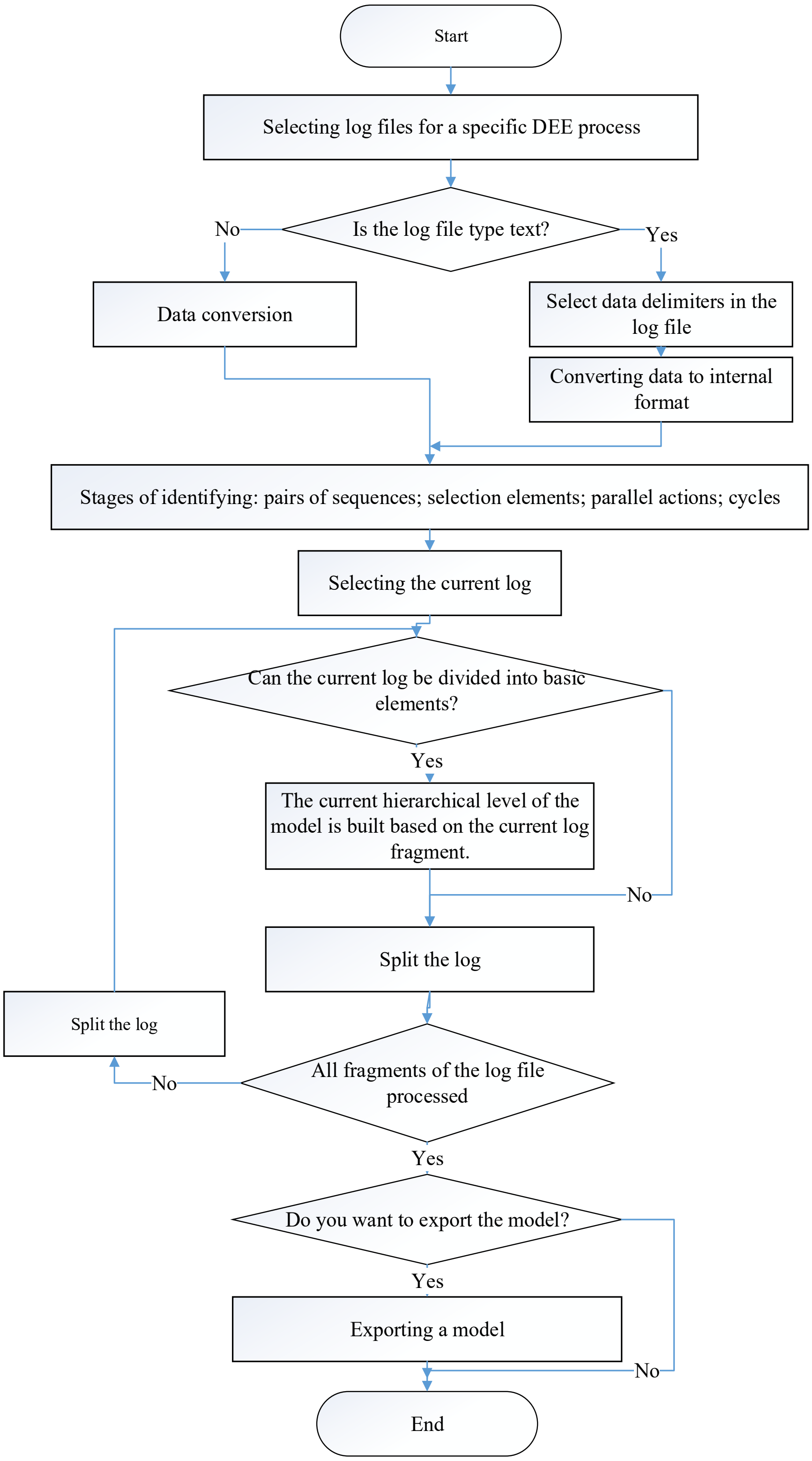


Fig. 2. Recursive algorithm for constructing the hierarchy of the analyzed process in the university's digital educational environment (DEE)

A distinctive feature of this algorithm is the recursive execution of stages during the construction of the hierarchy of the analyzed process in the university's Digital Educational Environment (DEE).

Having examined the recursive algorithm for constructing the hierarchy of the analyzed process of big data processing technology in the university's DEE, including personal data, we investigated the situation with the detection of unaccounted personal data in the DEE networks of the university. This was based on one of the possible ways of obtaining them – studying event registration logs.

Let's consider this case using the Splunk platform, which will help reveal hidden connections between logs and create a "profile of the university DEE user."

For the practical implementation of the set task, we took the log journals of the distance learning system (DLS) courses, containing access and secure logs. Upon examining the content of these files, it was found that among the thousands of lines of registered events, only a portion contains information. By analyzing this information, one can obtain data about the users of a specific DLS course. In our case, by "information," we mean: the time spent on the DLS site, the user's name (user_name), location (IP address), browser, operating system, the device used to view the DLS site pages, the list of links the course user clicked on, and more, as shown in Figure 3.

When examining these data separately from each other, we don't gain any valuable insights. However, if we have a complete "profile of the DLS user" with all this information, then it can be used for a detailed analysis of student behavior.

The practical implementation of the task of detecting unaccounted personal data was carried out using log files from several DLS courses of two universities - the National University of Bioresources and Nature Management of Ukraine and the Yessenov University (Kazakhstan). Event registration journals were analyzed for a period of 10 days at the beginning of the semester. By extracting all the IP addresses present in the logs, one can visualize the regions from which users access the DLS site. The query to visualize such a map in Splunk looks as follows:

*sourcetype=access_combined_wcookie | iplocation clientip | stats count by Country | geom geo_countries featureIdField="Country".*

Considering the fact that, for instance, during the COVID-19 pandemic, some students were located in regions other than their university's location, such information would be very useful for university management. For universities in Ukraine, which faced aggression from Russia, one can also use such or a similar query to view information about students who continue their studies while being abroad.

By tracking the activity of an entity over the time period for which we have data, one might obtain results such as: user activity under user_name: the number of events over a time span; the number of unsuccessful attempts to log into their DLS personal account; times of day when the user accesses the DLS site (for example, the user primarily accessed the DLS site in the morning); atypical cases of IP address changes; logins made from one or different devices, browsers, etc.

Such information about an individual user can be visualized, for example, as shown in Figure 4.

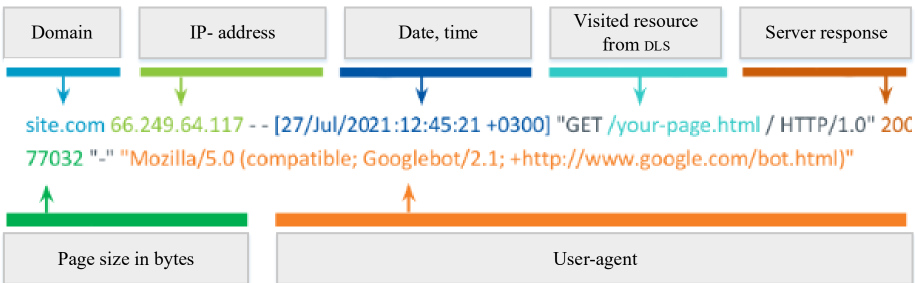


Fig. 3. Example of a log file structure from the University's Distance Learning System (DLS)

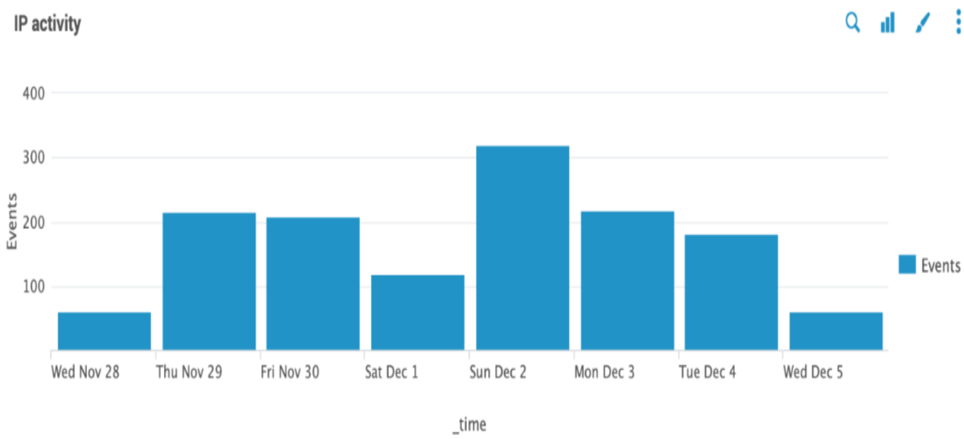


Fig. 4. Example of visualizing user activity data on the University's Distance Learning System (DLS) website in Splunk.

Having such information, the DLS administrator, as well as the management of the educational institution, can, for instance, evaluate the popularity (or relevance) of file resources in the DLS. When implementing such a task programmatically, one might use clustering analysis methods [16-20]. Such an assessment aids in more efficient management of user file resources, based on log file analytics, with the aim of identifying hidden patterns. These kinds of patterns can, in particular, assist in solving tasks related to optimizing the content of the DLS file server.

Let's assume, for example, that the conditional hierarchical log file of the DLS consists of the following parts: process; case; event. Of course, a more complex hierarchical structure can be used. We are considering, within the scope of this article, only a conceptual approach for using clustering analysis when processing data from DLS or university's Digital Educational Environment (DEE) log files.

By "process," we mean the aggregate of trajectories for processing all documents within the framework of a DLS educational discipline (i.e., the course syllabus, lectures, practical and laboratory work, joint educational projects, video content, news, chat, etc.). In essence, this is the document flow within the course.

A "case" represents the trajectory of the flow. Or a subprocess related to the processing of an identified course document. For example, a case corresponds to updating an answer to a practical assignment of the discipline or participating in the implementation of a joint educational project.

An "event" can be understood as a single operation of processing an electronic document in the university's DEE system.

An event is described in the format of a tuple $\langle doc, op, per \rangle$, where $doc-$ the corresponding document in the university's DEE, $op-$ the type of operation in the university's DEE, and $per-$ the individual who performed the operation (IPO).

Based on the described structure, let's consider the three components of the document turnover process in the DLS or the university's DEE: $Op \times Op$, $Per \times Per$ and $Per \times Op$. Then, for instance, the component $Op \times Op$ can be described by the $Q$ log file. This file corresponds to a set of sequential trajectories in the form $Q = \{tr_i | tr_i \in Op^*\}$. Here, the trajectory $tr-$ is a case for which there is such a sequence of operations that $tr \in Op^*$. $Op^*-$ is the set of all combinations from the alphabet of operation types with electronic documents (ED) in the DLS or the university's DEE - $Op$.

The second component describes the organizational structure of the ED turnover in the DLS or the university's DEE, i.e., $Per(Q)$. The second component can also be represented as a graph of collaborative work of IPOs. Such a graph $Gr$ will have $N$ vertices. Each vertex corresponds to a specific IPO, for instance, a teacher who conducts a course in the DLS or other IPOs registered in the university's DEE - $\{per_1, per_2, ..., per_N\}$.

The third component will depict the distribution of IPOs by operations of the corresponding processes.

Given that log files, as well as systems like Splunk, are often used for analyzing the state of information security (IS) of an object of information activity (OIA), which can include the university's DEE, below is an algorithm for identifying threats to the composition of IPOs that participate in the processes of ED processing. We used the clustering method.

We assume that the initial data will be:

- a set of IPOs $\{per_1, per_2, ..., per_N\}$ , which participate in the document flow process, for example, for specific courses in the DLS;
- a proximity matrix between actions $\{per_1, per_2, ..., per_N\}$ - $W_{ij}$ ;
- the composition of IPOs that participate in individual cases of ED processing - $per(doc)$ .

Collaborative work of students in small groups is a crucial component of effective learning. This is important because, in modern society, people spend a significant portion of their lives in small groups. In such groups, roles are typically distributed, and objectives are defined to address various tasks. The list of such tasks is extensive, ranging from social to production-related. In the latter case, relatively large teams work on solving complex tasks or entire problems. The primary objective in organizing student collaboration is to define common goals in the learning process. Effective student collaboration in education will allow them to learn and teach each other in "real-world" conditions. In traditional education, students complete assignments and receive corresponding individual grades. This learning model does not fully prepare them for professional activities in modern market conditions, where they must make decisions and work in teams to accomplish tasks too complex for individual resolution. In today's interconnected business world, to implement real projects, collaboration between companies and cooperation of people from different parts of the world is essential. This type of activity requires skills for effective collaboration, the ability to work productively in a team, and the integration of one's experience and ideas into a unified solution.

Below is a step-by-step algorithm, see Fig. 5, for identifying threats to the composition of IPOs participating in the implementation of a joint educational project, hosted, for example, in the DLS.

Step 1. Create $N$ groups with one IPO in each $c_1 = \{per_1\}, c_2 = \{per_2\}, \ldots, c_n = \{per_n\}.$;

Step 2. Calculate proximity metrics:

$$D(c_i, c_j) = \frac{1}{|c_i||c_j|} \cdot \sum_{x \le |c_i|} \sum_{y \le |c_j|} r_{xy}, \qquad (1)$$

Where $|c_i|-$ is the number of IPO vertices that belong to group $i$; $|c_j|-$ is the number of IPO vertices that belong to group $j$; $r_{xy}-$ represents the strength of the connection between the vertices of the graph.

Step 3. Merge the pair of groups characterized by the maximum proximity metric. The number of new groups will be $N' = N - 1$.

Step 4. Repeat steps 2, 3, and 4 until $N' = 1$.

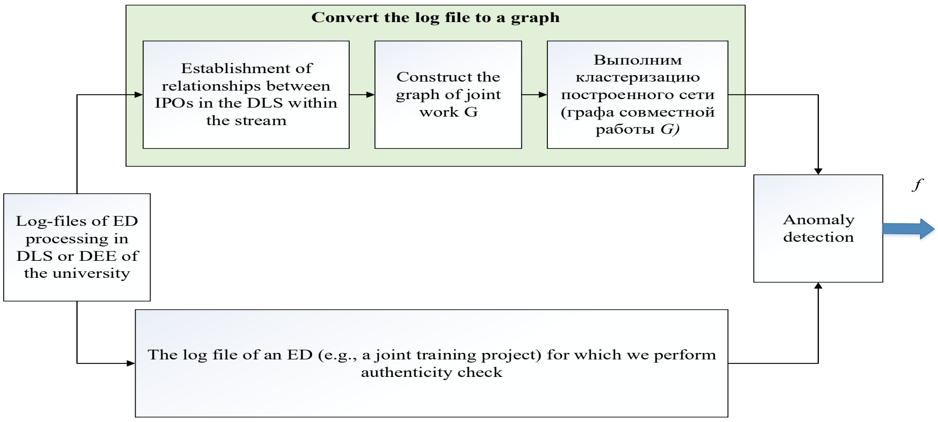


Fig. 5. Scheme for processing log files during the analysis of violations in the composition of participants in a collaborative educational project within the university's distance learning system

Step 5. Identify groups for clustering as $G = \{g_1, g_2, ..., g_{n-1}\}$.

Step 6. Calculate the modularity coefficients $V$ [21] for each group $g_i \in G$. For this, use the following formula [22]:

$$V = \left(1/(2 \cdot m)\right) \cdot \sum_{xy} \left(W_{xy} - \frac{k_x \cdot k_y}{2m}\right) \cdot \beta(u_x, u_y), \qquad (2)$$

Where $m-$ is the sum of the weights of the graph's edges; $W_{xy}-$ is the proximity matrix; $k_x, k_y -$ are the degree indicators of groups $x$ and $y$ (number of edges incident to them); $\beta(u_x, u_y) = 1$ if vertices $x$ and $y$ belong to the same group and 0 otherwise.

Step 7. Determine when the modularity coefficient $V$ [21] will have the maximum value. As a result, we obtain the best cluster division, $Per(Q) = \{Cl_1, Cl_2, ..., Cl_n\}$, where $Per(Q)-$ is the collective of students, for example, in a faculty; $Cl-$ represents the clusters.

Step 8. In the trusted log file database, find cases where the IPOs participating in the joint educational project are located in different clusters. For this, we look at the results obtained in the previous step. Accordingly, for $U_q = \{U_{q_1}, U_{q_2}, ..., U_{q_n}\}$, determine the proximity metrics for each set $U_q$ using formula (1).

Step 9. Determine the smallest proximity metric for the se $U_q = \{U_{q_1}, U_{q_2}, ..., U_{q_n}\}$t . Then, set the smallest proximity metric for the set as the boundary for identifying anomaly cases for $D_{\min}$ .

Step 10. Assume that the log file under analysis characterizes the set of IPOs participating in a joint educational project, i.e., $per(doc)$. Determine the smallest proximity metric value between $per(doc)$ and the clusters. The expression looks like this:

$$\min D(per(doc), K_i) =$$
$$= \min\left(D(per(doc), K_1), D(per(doc), K_2), ..., D((per(doc), K_i)\right).$$

Step 11. Determine the value of the function $f$ (shown by the blue arrow in Figure 5). This function indicates a possible anomaly of the IPO set participating in the document flow $per(doc)$. We assume that this document flow relates to $Per(Q)$ relatively.

The obtained results can be processed in conjunction with textual log files, for example, using a script in the R programming language.

## Conclusion

An algorithm has been proposed for constructing a process model in the Data Processing Center (DPC) based on the analysis of logs in the DPC. The algorithm allows describing a specific process in the DPC as a hierarchy of basic process elements. This algorithm takes into account the presence of additional data during the construction of the process hierarchy. The practical application area of the proposed algorithm is the methods of intellectual analysis of processes in the university's DPC.

It has been shown that the application of clustering methods can be useful for analyzing the registration logs of system processes in the university's Centralized Data Processing System (CDPS) or in the Distance Learning System (DLS). Such an analysis, for example, can help identify anomalous behavior of students and other individuals in the university's DPC. For instance, anomalous behavior can be considered a situation where people (IPOs) who have practically never participated in previous projects are involved in the processes of working on a joint educational project. Such a student behavior model, albeit indirectly, may indicate collusion. There is also a possibility of authorship substitution. Essentially, this is about academic plagiarism. As a result, the teacher faces the controversial authenticity of the educational project or joint work.

The algorithms proposed in the work allow, during the analysis of log files, to conduct research aimed at detecting violations of the information security of the university's DPC.

## REFERENCES


[1] Surmelioglu, Y., & Seferoglu, S. S. (2019). An Examination of Digital Footprint Awareness and Digital Experiences of Higher Education Students. World Journal on Educational Technology: Current Issues, 11(1), 48-64.

[2] McDermot, M. (2018). Digital footprints: Creation, implication, and higher education. FDLA Journal, 3(1), 11.

[3] Zhang, H. Z., Xie, C., & Nourian, S. (2018). Are their designs iterative or fixated? Investigating design patterns from student digital footprints in computer-aided design software. International Journal of Technology and Design Education, 28, 819-841.

[4] Songsom, N., Nilsook, P., Wannapiroon, P., Chun Che Fung, L. & Wong, K. (2019). System Architecture of a Student Relationship Management System using Internet of Things to collect Digital Footprint of Higher Education Institutions. International Journal of Emerging Technologies in Learning (iJET), 14(23), 125-140. Kassel, Germany: International Journal of Emerging Technology in Learning. Retrieved October 3, 2023 from https://www.learntechlib.org/p/217252/

[5] Abdulmohsen A. Data Mining in Education. (IJACSA) International Journal of Advanced Computer Science and Applications, 2016, vol. 7, no. 6, pp. 456-461.

[6] Angeli, C., Howard, S., Ma, J., Yang, J., Kirschner, P.A. Data mining in educational technology classroom research: can it make a contribution? Computers & Education, 2017. 113, pp. 226–242.

[7] Baker R., Yacef K. The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 2009, vol. 1, no. 1, pp. 3–17. https://doi.org/10.5281/ZENODO.3554657

[8] Bowers A. J., Sprott R., Taff S. A. Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. The High School Journal, 2012, vol. 96, no. 2, pp. 77–100. https://doi.org/10.1353/hsj.2013.0000

[9] Bowers A. J. Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. Practical Assessment Research and Evaluation, 2010, vol. 15. Article 7. https://doi.org/10.7275/r4zq-9c31

[10] Ezekiel U. O., Mogorosi, M. Educational Data Mining for Monitoring and Improving Academic Performance at University Levels. (IJACSA) International Journal of Advanced Computer Science and Applications, 2020, vol. 11, no. 11. pp. 570–591.

[11] Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. Expert systems with applications, 2014, vol. 41(4), pp. 1432-1462.

[12] R. Zamora-Musa and J. Velez. Use of Data Mining to Identify Trends between Variables to Improve Implementation of an Immer-sive Environment. Journal of Engineering and Applied Sciences, 2017, vol. 12, no. 22, pp. 5944-5948.

[13] Rawad C., Rémi B. Internationalizing Professional Development: Using Educational Data Mining to Analyze Learners' Performance and Dropouts in a French MOOC. International Review of Research in Open and Distributed Learning, 2020, vol. 21, no. 40.

[14] Razaque, Abdul & Alajlan, Abrar. Supervised Machine Learning Model-Based Approach for Performance Prediction of Students. Journal of Computer Science, 2020, 16, pp. 1150-1162.

[15] Siemens G., Baker R. S. J. d. Learning analytics and educational data mining: towards communication and collaboration / LAK'12. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. New York: Association for Computing Machinery, 2012, pp. 252–254. https://doi.org/10.1145/2330601.2330661

[16] Stauffer M. Laravel: Up & Running: A Framework for Building Modern PHP Apps 2nd Edition, O'REILLY, 2019.

[17] Rudometkina, M. N. (2014). Algoritm formirovaniya predikatnoy modeli gibkogo mnogovariantnogo protsessa na osnove analiza ego log-faylov. Fundamentalnyie issledovaniya, (11-1), 39-45.

[18] Prewett, J. E. (2003, June). Analyzing cluster log files using logsurfer. In Proceedings of the 4th Annual Conference on Linux Clusters. Citeseer.

[19] Chuah, E., Kuo, S. H., Hiew, P., Tjhi, W. C., Lee, G., Hammond, J., ... & Browne, J. C. (2010, December). Diagnosing the root-causes of failures from cluster log files. In 2010 International Conference on High Performance Computing (pp. 1-10). IEEE.

[20] Landauer, M., Wurzenberger, M., Skopik, F., Settanni, G., & Filzmoser, P. (2018). Dynamic log file analysis: An unsupervised cluster evolution approach for anomaly detection. computers & security, 79, 94-116.

[21] Goswami, A., & Polly, P. D. (2010). Methods for studying morphological integration and modularity. The Paleontological Society Papers, 16, 213-243.

[22] Tsibulya, A. N., & NGIA, H. M. (2016). Algoritm analiza zhurnala registratsii sistemnyih protsessov sistemyi elektronnogo dokumentooborota s ispolzovaniem metoda klasterizatsii sotsialьnyih setey. Informatsionnyie sistemyi i tehnologii, 94(2), 148.

[23] V. Lakhno, V. Malyukov, B. Akhmetov, B. Yagaliyeva, O. Kryvoruchko and A. Desiatko, "University Distributed Computer Network Vulnerability Assessment," 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST), Astana, Kazakhstan, 2023, pp. 141-144, https://doi.org/10.1109/SIST58284.2023.10223501

[24] Lakhno, V. *et al.* (2023). The Model of Server Virtualization System Protection in the Educational Institution Local Network. In: Shakya, S., Papakostas, G., Kamel, K.A. (eds) Mobile Computing and Sustainable Informatics. Lecture Notes on Data Engineering and Communications Technologies, vol 166. Springer, Singapore. https://doi.org/10.1007/978-981-99-0835-6_33