

*Е. В. Баранова, Н. О. Верещагина, Г. В. Швецов*

## ЦИФРОВЫЕ ИНСТРУМЕНТЫ ДЛЯ АНАЛИЗА УЧЕБНОЙ ДЕЯТЕЛЬНОСТИ СТУДЕНТОВ

*В статье исследуется проблема использования современных методов моделирования для анализа данных, формируемых при реализации образовательных процессов (Educational Data Mining — EDM). Описываются модели и цифровые инструменты, созданные авторами, предназначенные для мониторинга отчисления студентов, выявления направлений подготовки и образовательных программ, наиболее сложных для освоения, отслеживания динамики отчислений в различных разрезах: по формам обучения, укрупненным группам направлений и специальностей подготовки, отдельным направлениям и т. д. Рассматриваются подходы к выявлению методами EDM взаимосвязей между интенсивностью использования студентами цифровой образовательной среды и успешностью освоения дисциплин и модулей.*

**Ключевые слова:** анализ образовательной деятельности, цифровой след, цифровые ресурсы, цифровая образовательная среда, мониторинг освоения образовательных маршрутов.

*E. Baranova, N. Vereshchagina, G. Shvetsov*

## DIGITAL TOOLS FOR LEARNING ACTIVITIES ANALYSIS

*The article examines the use of modern modelling methods in analyzing data generated during the implementation of educational processes (Educational Data Mining — EDM). The paper describes the models and the digital tools the authors created to monitor expulsion rates, identify the fields of study and educational programs that pose most difficulties for students, and track the drop out dynamics in various sections: by forms of education, groups of areas and specialties, individual areas, etc. The authors analyze EDM approaches towards identifying the correlation between how frequently and efficiently students use the digital educational environment and how successful they are in learning academic disciplines and completing modules.*

**Keywords:** educational activity analysis, digital footprint, digital resources, digital learning environment, monitoring the development of educational paths.

Активное внедрение информационных технологий в различные сферы жизни, в том числе образовательный процесс, приводит к частичному или полному переносу деятельности человека в виртуальную среду. Информацию, размещаемую пользователем на своей странице в цифровой среде, можно рассматривать как продукт деятельности, который позволяет выделить особенности конкретного человека, строить предположения о его психологических особенностях и когнитивных способностях [14].

Цифровая среда на современном этапе является необходимым условием реализации учебного процесса в высшем профессиональном образовательном учреждении, следовательно, актуальным для образовательной практики является исследование взаимосвязи между контентом личного профиля студента и успешностью освоения им образовательного маршрута. В контексте таких исследований сформировалось понятие «цифровой след», которое можно определить, следуя [13], как массив данных о результатах

образовательной и проектной деятельности студента. Изучение цифрового следа студента позволяет осуществлять моделирование его когнитивных особенностей и использовать такую модель для прогнозирования успешности освоения образовательной программы. Цифровой след в образовательном пространстве — отчуждаемый результат образовательной деятельности человека. Все образовательное пространство, которое собирает цифровой след, должно быть особым образом спроектировано с применением педагогического дизайна [11].

Цифровой след студента начинает формироваться с момента его поступления в университет и содержит персональные данные, информацию о поступлении, направлении подготовки, образовательной программе, успеваемости в течение всего периода обучения, а также данные учебной аналитики, собираемые автоматически при работе студентов в системах дистанционного обучения и на платформах открытого образования при прохождении студентами онлайн-курсов. Согласно В. А. Ларионовой, А. А. Карасик, интеллектуальный анализ данных позволяет изучить поведенческие шаблоны студентов, узнать, как учесть индивидуальные особенности каждого при формировании образовательных траекторий [12].

Современное направление исследований, связанных с применением методов интеллектуального анализа данных, машинного обучения и статистики к разного рода информации об образовательном процессе, получило название анализ образовательных данных (EducationalDataMining) — АОД. В рамках этого направления разрабатываются и совершенствуются модели и методы обработки образовательных данных, направленные на совершенствование образовательного процесса [9].

Анализ зарубежных и отечественных источников [15–17; 20] позволяет выделить следующие направления использования методов АОД в образовательной практике, наиболее важные для высшего образования:

- моделирование поведения студентов в процессе обучения с целью прогнозирования развития их познавательных возможностей, выявления контингента студентов, входящих в группы «риска» с высокой вероятностью быть отчисленными;
- разработка новых моделей и способов представления знаний в предметной области, которые бы соответствовали разнообразным стилям обучения и познавательным возможностям, выявленным у студентов при помощи методов АОД;
- изучение процессов взаимодействия студентов с цифровой средой, выявление компонентов среды, изучения эффектов, которые оказывает образовательная среда на результаты этого обучения.

Одним из методов обработки образовательных данных является «перегонка данных для принятия решения человеком» (Distillation of Data for Human Judgment). Термин введен Д. Сименсом, Р. Бэйкером и предполагает кластеризацию данных, т. е. автоматическое разбиение элементов на группы по «схожести» набора характеристик с целью сокращения размерности исходного набора данных, и визуализацию данных, т. е. представление в виде, способствующем восприятию человеком [18].

Предметом данного исследования являются компоненты цифрового следа, представленные оценками, полученными студентами при освоении образовательных программ, в процессе промежуточной и итоговой аттестации, и деятельность студентов в СДО Moodle в процессе освоения дисциплин, а также взаимосвязь этих компонентов.

В РГПУ им. А. И. Герцена разработана и полномасштабно внедрена в практику образовательного процесса интегрированная информационная система управления учебным процессом — ИСУП (<http://oio.herzen.edu.ru>) [4; 5] для структурных подразделений, связанных с планированием, организацией

и управлением учебным процессом. ИСУП интегрирована в цифровую образовательную среду (ЦОС) университета и включает совокупность информационных систем (ИС) и веб-ресурсов, обеспечивающих в том числе формирование в распределенной базе данных (далее — БД HERZEN) целостной информации о различных аспектах образовательного процесса [1; 2; 7; 8].

Одним из важнейших компонентов ИСУП является ИС «Деканат», которая обеспечивает реализацию функционала, связанного с организацией и управлением образовательными маршрутами студентов [3; 6], включая формирование: персональных данных студентов, информации о характеристиках образовательной программы студента, информации о структуре и составе образовательной программы, данных об индивидуальных образовательных маршрутах студентов и успешности их освоения и т. д.

Приведем приблизительные количественные характеристики контента БД, сформированного более чем за 15 лет использования ИСУП в университете: количество укрупненных групп специальностей и направлений подготовки — более 30; направлений и специальностей подготовки — порядка 230; образовательных программ, в том числе за прошлые годы — более 6 000; дисциплин, модулей, практик — более 25 000; контингент студентов (действующие студенты, выпускники, отчисленные) — порядка 50 000 человек.

Приведенные числовые характеристики с очевидностью свидетельствуют о достаточно большом объеме данных, накопленных за длительный период. На основе методов перегонки данных были разработаны модели визуализации, наглядно представляющие процент отчисляемых из университета студентов в различных разрезах: по формам обучения, укрупненным группам направлений и специальностей подготовки (УГСН), отдельным направлениям и т. д.

На основе моделей разрабатывается цифровой ресурс, предназначенный для пред-

ставления руководителям образовательных подразделений, деканам факультетов, заведующим кафедрами, проректорам данных отчисляемых в процессе обучения студентов в разрезе уровней образования, форм обучения, направлений и специальностей подготовки. Получаемые данные позволят выявить на раннем этапе категории студентов, находящихся в группе «риска», образовательные программы и направления подготовки, сложные для освоения.

Ресурс является компонентом ИСУП, создается на базе современных технологий разработки веб-решений, в настоящий момент включает комплекс отчетов для руководства университета:

- 1) Отчеты для печати в формате PDF. Данный класс отчетов позволяет сформировать данные по определенным параметрам в табличном формате, а также вывести информацию на печать.
- 2) Детализированные отчеты. Такой класс отчетов позволяет ознакомиться с данными на различных уровнях детализации, которые определяются в соответствии со спецификой отчета; например, отчет о студентах предполагает следующие уровни: факультет/институт/филиал, студенты выбранного факультета, информация о конкретном студенте.
- 3) Отчет об отчисляемых студентах.

На первом уровне представляется обобщающая информация (рис. 1) о проценте отчисленных студентов (бюджет и внебюджет) выбранного уровня образования по формам обучения за указанный учебный год.

Проведенные расчеты за 2017–2020 уч. г. позволили сделать следующие выводы:

- наибольший процент отчисленных студентов приходится на 2017–2018 уч. г. и составляет 9,1 %;
- процент отчисленных по формам обучения различается следующим образом: наиболее высокий процент отчисленных студентов обучались на очно-заочной форме. Средние

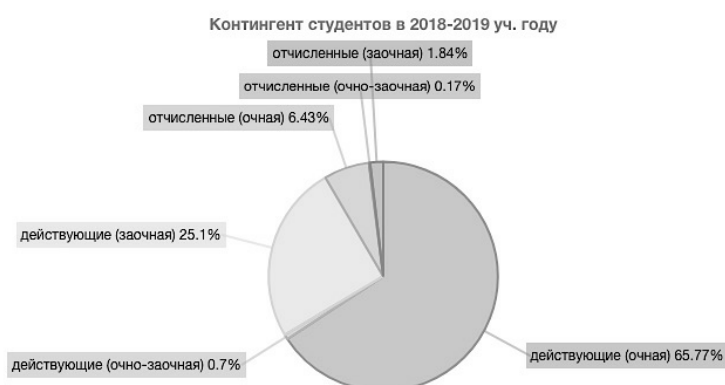


Таблица значений

№п/п	Форма обучения	Действующие	Отчисленные	Процент отчисленных
1	очная	12608	1232	8,9
2	очно-заочная	134	33	19,76
3	заочная	4811	353	6,84

Рис. 1. Контингент студентов по формам обучения

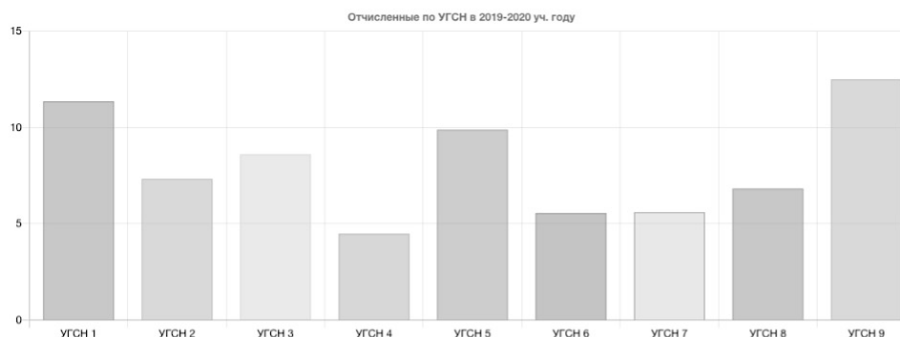


Таблица значений

№п/п	УГСН	Количество отчисленных студентов	Процент отчисленных студентов
1	УГСН 1	5	11,36
2	УГСН 2	14	7,33
3	УГСН 3	44	8,61
4	УГСН 4	3	4,48
5	УГСН 5	9	9,89
6	УГСН 6	6	5,56
7	УГСН 7	158	5,6
8	УГСН 8	11	6,83
9	УГСН 9	10	12,5

Рис. 2. Отчисленные студенты по УГСН

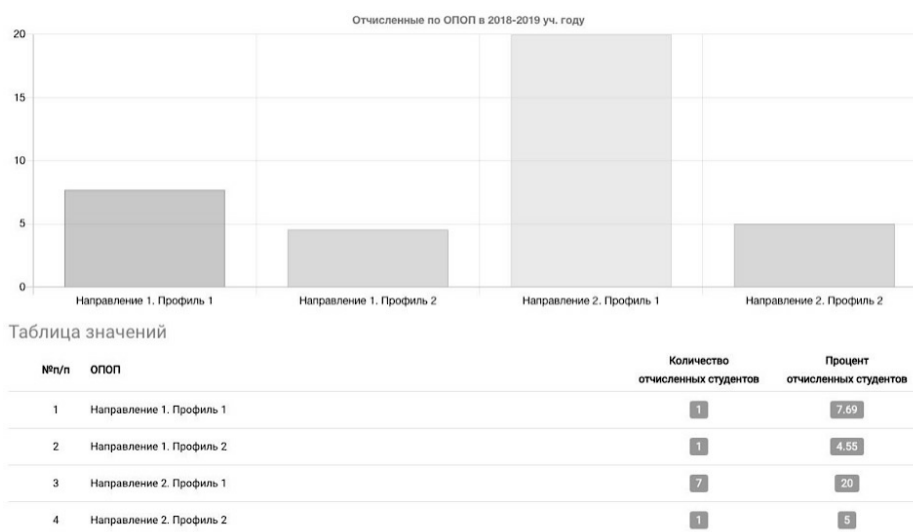


Рис. 3. Отчисленные студенты по образовательным программам

показатели отчисленных студентов по формам обучения составляют: очная — 8,62 %, очно-заочная — 19 %, заочная — 6,96 %;

- анализ данных показал, что за расчетный период заметно уменьшение доли отчисленных студентов.

Для каждой формы обучения можно получить данные об отчисленных студентах по УГНС (бюджет и внебюджет) за выбранный учебный год (рис. 2), выявить группы направлений с наибольшим количеством отчисленных.

По выбранному направлению можно получить данные об отчисленных студентах по образовательным программам (рис. 3).

Использование разработанного цифрового инструмента позволяет выявить образовательные программы, которые сложнее осваиваются студентами, а также отследить динамику движения контингента по годам.

Дальнейшее уточнение модели позволят выявлять «критические» для студентов этапы освоения образовательных маршрутов (курсы, дисциплины) в зависимости от направлений подготовки, образовательных программ и т. д.

Разработанные инструменты позволяют осуществлять мониторинг освоения студентами образовательных маршрутов, выявлять наиболее сложные этапы, принимать меры и, следовательно, повышать эффективность и качество обучения.

Вторая группа моделей и цифровых инструментов направлена на получение данных, выявляющих зависимость между оценками студентов по промежуточной успеваемости и интенсивностью их работы в цифровой образовательной среде.

Вопросам определения показателей, которые характеризуют интенсивность использования студентами цифровой образовательной среды, посвящено много работ как отечественных, так и зарубежных исследователей. В основном предлагается оценивать частоту и длительность посещений студентами компонентов ресурса. В частности, К. Ромеро и С. Вентура предлагают следующие группы показателей:

- число посещений и число посетителей страниц ресурса, длительность посещений за определенный период времени;
- число различных страниц, просмотренных или скачанных студентом

за сеанс работы или за более продолжительный период времени;

- статистические индикаторы общения на форуме СДО (число постов, средний объем поста, популярность поднятой темы, количество обращений с вопросами к другим студентам по сравнению с общением «студент — преподаватель») и т. д. [19].

Современные условия определяют необходимость интеграции различных технологий дистанционного обучения и традиционных форм реализации учебного процесса. В этих условиях актуальной представляется задача выявления взаимосвязи между интенсивностью использования студентами цифровых образовательных ресурсов и успешностью освоения образовательных программ в целом.

В университете дисциплины по всем образовательным программам подготовки реализуется с использованием электронных учебных курсов (ЭУК), размещенных на платформе СДО Moodle. Обучающая среда Moodle является на сегодняшний день одной из наиболее популярных систем поддержки учебного процесса в условиях дистанционного образования. Важнейшими преимуществами среды, обеспечивающими ее широкую востребованность, являются: открытость, мобильность, переносимость, расширяемость, широкая распространенность и т. д.

СДО Moodle содержит широкий набор инструментов для создания и сопровождения электронных учебных курсов, различные возможности для представления учебного материала, контроля успеваемости, реализации обратной связи между педагогом и студентами, которая позволяет студентам выкладывать работы в СДО, получать рецензии и консультации преподавателя [10].

Простые по структуре учебные курсы содержат учебно-методические материалы для освоения дисциплины в виде презентаций, текстов лекций, литературы, вопросов к экзамену и т. д. Большая часть курсов, размещенных в СДО Moodle университета, предполагает более широкое использование

различных инструментов СДО и включает, как правило: рабочую программу дисциплины; учебно-методические материалы для освоения дисциплины; методические рекомендации для студентов по изучению ЭУК и подготовке к различным видам занятий, аттестации; форум для общения и обсуждения студентами и преподавателем вопросов по курсу; фонды оценочных средств для контроля знаний студентов.

СДО Moodle, базирующаяся на СУБД MariaDB (ответвление СУБД MySQL), обеспечивает хранение различных данных о взаимодействии студентов с цифровой средой в процессе обучения. Аппарат запросов к БД Moodle позволяет получить различные наборы данных, включая:

- общие данные о курсе, преподавателях и студентах;
- количество проведенного студентом в онлайн-курсе времени;
- количество изменений, произведенных студентами в различных модулях курса;
- результаты выполненных заданий студентов на курсе;
- информацию об активности студентов на форумах курса и т. д.

Исследование взаимосвязи между интенсивностью работы студентов в СДО Moodle и успешностью их обучения осуществлялось в несколько этапов.

На первом этапе были выделены три показателя, значения которых могут быть рассчитаны автоматически с использованием запросов к БД MariaDB:

- 1) количество действий, выполненных с элементами курса (загрузка ресурсов, просмотр лекций и т. д.);
- 2) активность студентов в чате курса;
- 3) итоговая оценка за курс.

На следующем этапе был выбран ряд дисциплин, играющих ключевую роль в подготовке по различным образовательным программам, освоение которых базируется на широком использовании ЭУКов в СДО Moodle: геометрия, теория организации

и управление организационной политикой в образовании, систематика растений, история Древнего мира, защита и сохранение культурного наследия и др.

Информация об успешности освоения студентами образовательной программы формируется в виде оценок за промежуточную аттестацию по модулям, дисциплинам и практикам в БД Herzen. Для связи данных из двух баз была разработана модель интеграции (рис. 4) и осуществлена программная реализация.

Исследование базировалось на данных об обучении 5 групп обучающихся, среднее количество студентов в которых составляет 20 человек. Для каждой дисциплины, группы студентов и выбранному показателю были рассчитаны коэффициенты корреляции, отражающие зависимость значений показателя

и оценки студентов в рамках промежуточной аттестации по дисциплинам (табл. 1).

Показатели количества действий студентов и их активности в чате(-ах) курса характеризуют побочную деятельность студентов для достижения конечного результата, значение итоговой оценки за курс формируется на основе качества выполнения заданий, предложенных студентам в процессе прохождения онлайн-курса.

Приведенные результаты позволили сделать следующие выводы:

- наибольший средний коэффициент корреляции соответствует показателю итоговой оценки студента за курс — 0,6286; согласно шкале Чеддока, такая количественная мера тесноты связи характеризуется как заметная;

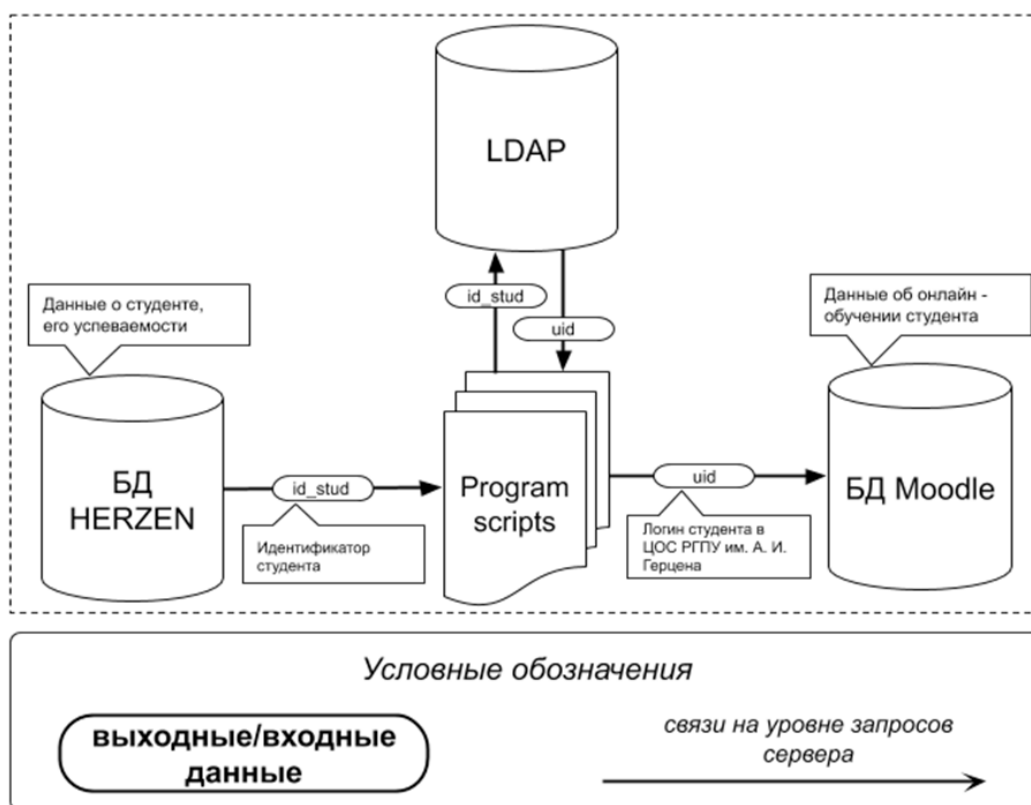


Рис. 4. Модель интеграции БД HERZEN и БД MariaDB

Таблица 1

## Коэффициенты корреляции (r) для выбранных показателей

Дисциплина	Кол-во студентов в группе	Значения r для показателя 1	Значения r для показателя 2	Значения r для показателя 3
Геометрия	20	0,2582	0,051	0,4014
Теория организации и управление организационной политикой в образовании	27	0,6349	0,0308	0,917
Систематика растений	16	0,1388	0,4070	0,9369
История Древнего мира	29	0,1935	0,0494	0,6303
Защита и сохранение культурного наследия	27	0,3447	0,5205	0,6959

- наименьший средний коэффициент корреляции соответствует показателю активности студентов в чате(-ах) курса — 0,2117, такая связь является слабой.

Предложенный авторами подход ориентирован на выявление наиболее значимых для результатов освоения дисциплин показателей и создает базу для дальнейших исследований.

Разработанные авторами модели, базирующиеся на методах анализа образовательной деятельности, и цифровые инструменты позволяют осуществлять мониторинг успешности освоения студентами образовательных маршрутов, а также исследовать взаимосвязь между интенсивностью учебной деятельности студента в цифровой образовательной

среде и результатами промежуточной успеваемости. Развитие предложенных подходов позволит в дальнейшем:

- выявлять на раннем этапе категории студентов, находящихся в группе «риска», образовательные программы и направления подготовки, сложные для освоения, «точки» образовательных маршрутов (курсы, дисциплины), которые являются препятствием для студентов;
- исследовать взаимодействия студентов с цифровой средой с целью выявления компонентов среды, в наибольшей степени используемых студентами и способствующих усвоению учебного материала.



### СПИСОК ЛИТЕРАТУРЫ

1. Баранова Е. В. Современная информационно-образовательная среда вуза как механизм реализации требований стандартов нового поколения // Известия Российского государственного педагогического университета им. А. И. Герцена. 2015. № 177. С. 70–73.
2. Баранова Е. В., Верещагина Н. О. Инновационные ресурсы для организации учебного процесса в современном вузе в условиях цифрового образования // Письма в Эмиссия. Офлайн: электронный научный журнал. 2018. № 10. С. 2657.
3. Баранова Е. В., Гизатуллина Г. С. Модель веб-ресурса «Деканат» как компонента интегрированной системы управления учебным процессом // Сборник научных статей по материалам международной научной конференции 1–12 апреля 2019 года. СПб.: Изд-во РГПУ им. А. И. Герцена, 2019. С. 144.
4. Баранова Е. В., Елизарова И. К. Интегрированная информационная система управления образовательными маршрутами в вузе («Электронный атлас»). Свидетельство о государственной регистрации программы для ЭВМ № 2011613872. 18.05.2011 г.
5. Баранова Е. В., Елизарова И. К., Верещагина Н. О. и др. Программный комплекс для управления учебным процессом в вузе (ПК «Герцен»). Свидетельство о государственной регистрации программы для ЭВМ № 2019618795. 04.07.2019 г.
6. Баранова Е. В., Елизарова И. К., Слепухина Н. В. Информационная система «Деканат». Свидетельство о государственной регистрации программы для ЭВМ № 2014616013. 09.06.2014 г.
7. Баранова Е. В., Симонова И. В. Модели ресурсов электронной информационно-образовательной среды для решения профессиональных задач преподавателя педагогического вуза // Информатика и образование. 2016. № 9. С. 18–21.
8. Баранова Е. В., Швецов Г. В. Цифровая модель индивидуальных образовательных маршрутов студентов // Новые образовательные стратегии в современном информационном пространстве: сборник научных статей по материалам международной научной конференции. Санкт-Петербург, 9–25 марта 2020 года. СПб.: Изд-во РГПУ им. А. И. Герцена, 2020. С. 129–133.
9. Белоножко П. П., Карпенко А. П., Храмов Д. А. Анализ образовательных данных: направления и перспективы применения // Интернет-журнал «Науковедение». 2017. Том 9. № 4. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/analiz-obrazovatelnyh-dannyh-napravleniya-i-perspektivy-primeneniya/viewer> (дата обращения: 05.07.2020).
10. Зайцева О. Н. Использование LMS Moodle в образовании // Обучение и воспитание: методики и практика. 2012. № 2. С. 59–64.
11. Курбацкий В. Н. Цифровой след в образовательном пространстве как основа трансформации современного университета // «Высшая школа»: научно-методический и публицистический журнал. 2019. № 5. С. 40–45.
12. Ларионова В. А., Карасик А. А. Цифровая трансформация университетов: заметки о глобальной конференции по технологиям в образовании Edcrunch Ural // Университетское управление: практика и анализ. 2019. № 23. С. 130–135.
13. Степаненко А. А., Феценко А. В. «Цифровой след» студента: поиск, анализ, интерпретация // Открытое и дистанционное образование. 2017. № 4 (68). С. 58–62. DOI: 10.17223/16095944/68/9
14. Тулупьева Т. В., Суворова А. В., Азаров А. А., Тулупьев А. Л., Бордовская Н. В. Возможности и опыт применения компьютерных инструментов в анализе цифровых следов студентов-пользователей социальной сети // Компьютерные инструменты в образовании. 2015. № 5. С. 3–13.
15. 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. P. 3–17. DOI: 10.5281/ZENODO.3554657
16. 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. DOI: 10.7275/r4zq-9c31
17. 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. P. 77–100. DOI: 10.1353/hsj.2013.0000
18. Siemens G., Baker R. S. J. d. 2012. Learning analytics and educational data mining: towards communication and collaboration // LAK'12. Proceedings of the 2<sup>nd</sup> International Conference on Learning Analytics and Knowledge. New York: Association for Computing Machinery, 2012. P. 252–254. DOI: 10.1145/2330601.2330661

19. *Romero C., Ventura S.* Educational data mining: A review of the state of the art // IEEE Transactions on Systems Man and Cybernetics. Part C (Applications and Reviews). 2010. Vol. 40. No. 6. P. 601–618. DOI: 10.1109/TSMCC.2010.2053532.

20. *Vaval L., Bowers A. J., Snodgrass Rangel V.* Identifying a typology of high schools based on their orientation toward STEM: A latent class analysis of HSLs:09 // Science Education. 2019. Vol. 103. No. 5. P. 1151–1175. DOI: 10.1002/sc.21534