

# К вопросу индивидуализации образовательных траекторий при управлении контролем знаний (на основе Big Data)

Н. А. Лойко<sup>1</sup>, А. Б. Петроченков, А. Г. Лейсле

*Пермский национальный исследовательский политехнический университет*

<sup>1</sup>E-mail: nataly.loyko@yandex.ru

**Аннотация.** В технических системах алгоритмы Big Data могут быть реализованы в полностью автоматическом режиме, без какого-либо участия человека. В социальных (образовательных, человеко-машинных) системах алгоритмы предупредительных и корректирующих действий невозможны без человеческого участия. Рассмотрен вопрос индивидуализации образовательных траекторий по результатам входного и рубежного тестирования, а это, в свою очередь, требует наличия идеального уровня качества баз тестовых заданий как главного компонента в системе педагогических измерений.

**Ключевые слова:** индивидуализация образовательных траекторий; входное и рубежное компьютерное тестирование; контроль знаний; качество баз тестовых заданий; управление на основе данных

## I. ВВЕДЕНИЕ

Сегодня, по прошествии 15 лет после исторического выпуска журнала «Nature» [1], посвященного презентации того, что в последствии было названо «Большими Данными», можно с уверенностью констатировать два обстоятельства:

1. Публикации с термином «Big Data» в заголовках можно разделить на две категории – во-первых, отчеты о реальных внедрениях законченных разработок, приносящих значимый положительный эффект, и, во-вторых – «новости будущего времени», то есть описания перспектив и потенциальной выгоды от еще несостоявшихся достижений. Несложный анализ показывает, что первая категория относится к так называемой «цифровой сфере» [2], охватывающей все то, что либо изначально существует в цифровой форме, либо с легкостью может быть подвержено дематериализации («оцифрованию»), и оперирующей почти исключительно виртуальными сущностями – например, электронная торговля, сотовая связь, геоинформационные системы и пр.

Вторая категория связана с так называемой «материальной сферой», в которой также циркулируют огромные объемы информации, но в силу разных причин дематериализация, отрыв от источника этой информации крайне затруднен [3]. Примером материальной сферы является система высшего профессионального образования, информационный обмен в которой производится, как правило, в вербальной форме, поскольку учебно-воспитательный процесс является не столько информационным, сколько психофизиологическим.

2. Другим, совершенно аналогичным по результатам разделения, обстоятельством, является степень возможной автоматизации анализа данных в указанных сферах. Поиск закономерностей даже в самых больших наборах данных в цифровой сфере не требует дополнительного изобретательства, и сводится к стандартному набору инструментов статистического, факторного и кластерного анализа. Стоящий особняком нейросетевой анализ также позволяет аналитикам использовать стандартный инструментарий сторонних разработчиков.

В материальной же сфере роль человека – эксперта в решении аналитических задач является ключевой. Экспертные заключения, конечно, также делаются с привлечением указанных выше инструментов, но используемых лишь на уровне предварительной, черновой обработки данных, сгенерированных той или иной социальной системой. Вычислительные алгоритмы в материальной сфере интегрируют, представляют в более наглядной форме массивы информации, понижая их размерность и делая возможным дальнейший, основанный на жизненном опыте, здравом смысле и профессиональной интуиции, процесс формирования тех или иных управленческих решений [4].

Принципиальное различие в характере цифровой и материальной сфер определяет различие количественного значения критерия, позволяющего относить (или не относить) информационные наборы, характерные для каждой из них, к категории «Big Data». Если в цифровой сфере «Большими Данными» являются массивы информации, которые невозможно хранить и обрабатывать на отдельно взятой вычислительной машине, то в материальной сфере таковыми являются информационные наборы, неподвластные уяснению отдельно взятым экспертом. При этом необходимость привлечения к анализу в материальной сфере команды экспертов совершенно аналогично массово – параллельной обработке «Big Data» в цифровой сфере.

## II. ПОСТРОЕНИЕ ИНДИВИДУАЛЬНЫХ ОБРАЗОВАТЕЛЬНЫХ ТРАЕКТОРИЙ ПО РЕЗУЛЬТАТАМ КОМПЬЮТЕРНОГО ТЕСТИРОВАНИЯ

В образовательной сфере одним из процессов, порождающих значительные объемы дематериализованной информации, является контроль учебных достижений обучаемых, и, в первую очередь, системы компьютерного тестирования. В информационно-аналитической системе Пермского национального исследовательского политехнического университета (ПНИПУ) за последние семь лет

накоплены данные о результатах тестирования 818 учебных групп, прошедших, в общей сложности, 16000 компьютерных тестирования. Учебные планы ПНИПУ предусматривают для большинства групп тестирование по 8 модулям учебных дисциплин «Математика» и «Физика», и по 3 модулям дисциплин «История» и «Философия». В тестировании приняли участие 19000 студентов, которые посетили компьютерные классы в общей сложности 164852 раз. Суммарно все студенты выполнили 3500000 тестовых задания, получив при этом 57632 оценки «неудовлетворительно», 56691 оценку «удовлетворительно», 33601 оценку «хорошо» и 16928 оценок «отлично».

Конечно, по меркам цифровой сферы подобная статистика выглядит микроскопичной. Но с точки зрения материальной сферы она вполне соответствует приведенному выше критерию, так как принятие всех необходимых управленческих решений хотя бы одного типа не под силу отдельно взятому эксперту. Проиллюстрируем данный тезис следующим примером.

Результаты компьютерного тестирования используются регулярно – для принятия решений о допуске студента к экзамену по той или иной учебной дисциплине, о возможном получении итоговой оценки по результатам тестирования, о систематическом невыполнении студентом графика учебного процесса, и т. д. Наибольшую важность эти результаты приобретают в ходе решения задач персонализации учебной работы и разработки индивидуальных образовательных траекторий. Проблема состоит в том, что уровень школьной подготовки абитуриентов, поступающих на разные факультеты, направления подготовки и специальности, существенно различен, и зависит от их популярности у поступающих в ВУЗ. Объем остаточных знаний первокурсников горно-нефтяного факультета почти вдвое превышает аналогичный показатель первокурсников механико-технологического факультета. Объем остаточных знаний первокурсников специальности «Разработка месторождений полезных ископаемых» не превышает двух третей от остаточных знаний первокурсников специальности «Прикладная геодезия». Студент группы МС-22-26 И.И. Иванов успешно выполнил 3 задания входного теста по «Математике», а студент группы ПГ-22-16 С.С. Сидоров – 19 заданий.

Проблема дефицита кадров на предприятиях машиностроения не позволяет отказывать в поступлении абитуриентам даже с недопустимо низкой школьной подготовкой. Для этих студентов организуются дополнительные занятия по всем разделам и темам школьной программы «Математики» и «Физики» с целью «дотянуть» их до минимально необходимого уровня понимания изучаемого материала, что требует от ПНИПУ немалых усилий со стороны преподавателей, значительных затрат на оплату их труда, выделения крайне дефицитного аудиторного фонда и т. д. Поэтому очень остро стоит задача оптимизации объема дополнительных занятий за счет их индивидуализации, в результате которой каждый отдельный студент должен посетить лишь те из них, которые связаны с конкретными «пробелами», показанными им на входном тестировании.

На рис. 1 приведена процедура формирования индивидуальных образовательных траекторий по результатам входного компьютерного тестирования

студентов. Для этого информационно-аналитическая система ПНИПУ содержит базы тестовых заданий, в первую очередь, по основным для технического ВУЗа учебным дисциплинам – «Математика» и «Физика», причем структура баз тестовых заданий соответствует структуре основных дидактических единиц (разделов, тем и понятий) школьной программы. Процедура входного тестирования производится в факультетских компьютерных классах ПНИПУ под управлением квалифицированных тестологов, решающих, в частности, проблему исключения несанкционированного доступа студентов к источникам информации, в первую очередь – в сети Интернет.

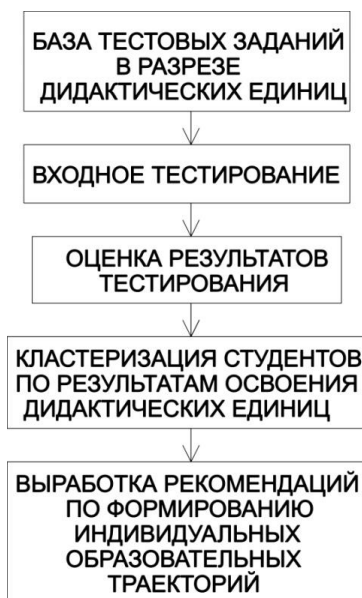


Рис. 1. Процедура формирования индивидуальных образовательных траекторий по результатам входного компьютерного тестирования

Оценка результатов входного тестирования производится также в разрезе таких основных понятий тезауруса учебных дисциплин школьной программы, твердое освоение которых необходимо для изучения тех или иных разделов и тем уже соответствующего вузовского курса.

Для обеспечения требуемого методического разнообразия в университете для учебных дисциплин «Математика» и «Физика» разработаны рабочие программы, характеризующихся разными наборами изучаемых разделов, в зависимости от требований направлений подготовки, а также темпом учебной работы. Это связано с тем, что разные профили одного и того же направления подготовки пользуются разной популярностью у абитуриентов, что приводит к разному уровню образования студентов, зачисленных в разные учебные группы. Поэтому изучение одних и тех же разделов одной учебной дисциплины в группах с более низким уровнем подготовленности требует большего учебного времени, чем в группах с хорошо подготовленными студентами. Для этого, в частности, изучение «Математики» в ПНИПУ предусматривает три варианта трудоемкости – 12, 16 и 21 зачетных единиц.

Кроме результатов входного тестирования, индивидуализация образовательного пространства определяется результатами рубежного компьютерного тестирования. Несмотря на то, что в современных

образовательных стандартах отсутствует требование обязательности модульного подхода, учебный процесс в ПНИПУ предусматривает централизованное компьютерное тестирование по окончании изучения разделов основных учебных дисциплин – «Математика», «Физика», «История», «Философия» и «Информатика». Основным результатом такого рубежного контроля является подробная информация, предоставляемая каждому преподавателю указанных дисциплин, о том, с какими именно тестовыми заданиями, и как именно, не справился тот или иной студент. Наряду с интегральной оценкой, такие сведения составляют основу индивидуального подхода к обучению каждого студента. Кроме того, эти сведения, интегрированные по учебным студенческим группам, специальностям и направлениям подготовки, обеспечивают прогностическую функцию рубежного контроля, позволяя по уже имеющимся результатам предвидеть последующие, и выбрать соответствующий вариант преподавания дисциплины по трудоемкости на будущее.

Преподаватели получают объективную информацию о том, насколько подробно и обстоятельно следует доносить до студентов материал того или иного раздела на аудиторных занятиях, а также об объеме и содержании самостоятельной работы, с которыми в состоянии справиться данный студенческий коллектив. Кроме того, рубежный контроль указывает преподавателям на допущенные ими (вольно или невольно) методические или дидактические ошибки, проявляющиеся в том, что некоторые темы или понятия усвоены необъяснимо (другими причинами) малым числом студентов. Очевидно, что процедура исправления каждой подобной ошибки не может предусматриваться типовой рабочей программой соответствующей учебной дисциплины, и носит индивидуальный характер.

### III. ВОПРОСЫ КАЧЕСТВА БАЗ ТЕСТОВЫХ ЗАДАНИЙ

В соответствии с устоявшимися требованиями, базы тестовых заданий должны соответствовать требованиям необходимого уровня сложности, дифференцирующей способности, ретестовой надежности, эквивалентности, валидности и пр. Существуют методики, позволяющие по уже накопленным результатам предыдущих тестов автоматически вычислять численные значения указанных уровней для каждого задания, входящего в состав той или иной базы тестовых заданий, что, на первый взгляд, позволяет отнести систему компьютерного тестирования к указанной выше «цифровой сфере». В действительности же это далеко не так.

Ключевая проблема оценки и управления сложностью баз тестовых заданий состоит в том, что фактическое значение апостериорной «статистической» сложности того или иного тестового задания становится известным только после того, как данное тестовое задание оказывается предъявленным не менее нескольких десятков раз. Однако для того, чтобы накопить подобную статистику, база тестовых заданий уже должна быть не просто разработана, но и эксплуатироваться в течение достаточно долгого времени. Очевидно, что задействование такого сложного и дорогого механизма, каким является компьютерное тестирование, в «холостом» режиме (без выставления оценки испытуемым) совершенно не возможно. Поэтому

в процессе составления тестовых заданий необходимо осуществлять априорную, обычно экспертную, оценку их сложности. Практика показывает, что данная априорная оценка сложности тестовых заданий зачастую не совпадает с более объективной апостериорной оценкой. На рис. 2 приведены результаты входного компьютерного тестирования студентов ПНИПУ по учебной дисциплине «Физика» в 2016–2021 годах. По оси абсцисс выстроены тестовые задания (ТЗ), отнесенные к трем категориям сложности, упорядоченные по возрастанию доли правильных ответов на них.

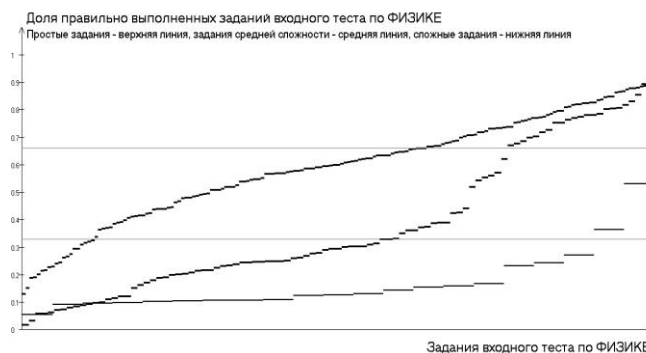


Рис. 2. Апостериорная оценка сложности ТЗ по «Физике» за 2016–2021 годы

Анализ 107684-х ответов студентов показал, что лишь задания, априорно отнесенные к категории «сложных», действительно являются таковыми. В то же время около половины априорно «простых» заданий апостериорно следует отнести к категории «средней сложности» (в соответствии с условными границами в виде тонких линий на рисунке). Эти, а также и другие необходимые изменения апостериорной оценки сложности ТЗ были внесены при подготовке к входному тестированию осенью 2022 года, что объективно повысило его результаты.

Сложность (как и трудность) является наиболее органичным свойством тестовых заданий, которое очень просто оценить, хотя и непросто изменить. Вместе с тем, достижение того или иного уровня сложности отнюдь не является самоцелью проектирования баз тестовых заданий ни при одном из видов тестирования.

Критериально-ориентированный тест не ставит своей задачей разделить испытуемых по уровню достижений, ограничившись определением того, освоили ли испытуемые требуемый, заранее определенный, объем знаний и/или умений. Требования к широте и глубине той части предметной области, факт освоения или не освоения, которой должен быть подтвержден тестированием, формируются в большей степени, исходя из организационно-административных, чем из педагогических соображений. Поэтому считается вполне нормальным, если с критериально-ориентированным тестом успешно справятся (либо, наоборот, не справятся) все до единого испытуемые. Конечно, фактор сложности и в этом случае является ключевым, но его проявление скрыто тем обстоятельством, что критериально-ориентированный тест редко бывает одноэтапным и даже одношаговым. Успешно справившись с первым, начальным, этапом испытаний, испытуемый, иногда немедленно, а зачастую спустя значительное время, пытается доказать факт освоения более сложного

участка предметной области. Однако среди тех, кто успешно справился с тестом, нет лучших или худших, все они «прошли этап», а с результатом 76 % или 84 % – не имеет значения, по крайней мере, формально.

Дифференцирующей способностью (дискриминативностью) теста называют его способность разделять испытуемых на более или менее успешных («лучших» и «худших»). Очевидно, что уровень дискриминативности теста определяется содержанием баз тестовых заданий, лежащих в его основе. В свою очередь, поскольку качество баз тестовых заданий зависит от качества содержания составляющих ее заданий, то следует говорить о дискриминативности каждого отдельного задания (в этом случае, используют термин «дифференцирующая сила»). На рис. 3 приведен результат более глубокого анализа ответов абитуриентов на задания входного теста по «Физике».

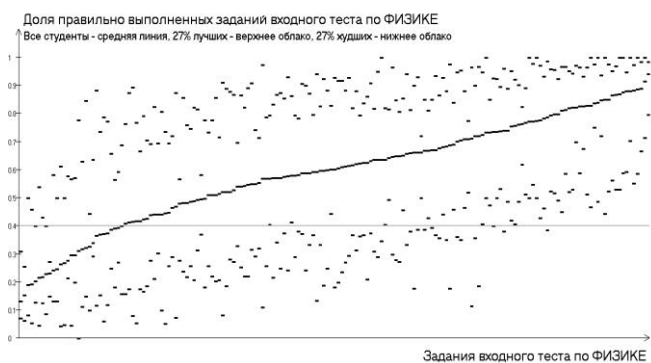


Рис. 3. Результаты компьютерного тестирования студентов ПНИПУ по «Физике» (27-процентные крайние группы)

Как и ранее, вдоль оси абсцисс расположены все тестовые задания модуля в порядке убывания их сложности (иначе – возрастания доли студентов, успешно справившихся с каждым из них). Средняя линия соответствует доле правильных ответов всех студентов ПНИПУ, протестированных по данному модулю, верхнее облако – результатам 27 % студентов, получивших наиболее высокие оценки за тест в целом, а нижнее облако – результатам 27 % самых слабых студентов. Как видно из рисунка, наибольшей дискриминативностью обладают ТЗ, соответствующие средним значениям уровня сложности, – в этом случае результаты так называемых «крайних групп» максимально разнесены. У самых простых и самых сложных заданий дифференцирующая способность недопустимо низкая, т. к. с ними одинаково успешно (или одинаково неуспешно) справляются как «сильные», так и «слабые» студенты.

Для численного выражения дифференцирующей силы ТЗ обычно используют индекс дискриминации, отражающий разную вероятность успешного выполнения того или иного ТЗ представителями верхней и нижней крайних групп:

$$D_i = (P_{\max})_i - (P_{\min})_i \quad (1)$$

где  $D_i$  – коэффициент дискриминативности  $i$ -го ТЗ;  $(P_{\max})_i$  – вероятность успешного выполнения  $i$ -го ТЗ представителем группы наиболее успешных студентов;  $(P_{\min})_i$  – вероятность успешного выполнения  $i$ -го ТЗ представителем группы наименее успешных студентов.

В дальнейшем, во избежание чрезмерной громоздкости, будем опускать индекс  $i$ . Выражение (1) справедливо лишь в отношении гипотетической генеральной выборки бесконечного объема. Для практического использования необходимо перейти от вероятностей к относительным частотам:

$$D = \frac{N_{n_{\max}}}{N_{n_{\max}}} - \frac{N_{n_{\min}}}{N_{n_{\min}}}$$

где  $N_{n_{\min}} = N_{n_{\max}}$  – общая численность испытуемых в крайних группах, которым в процессе тестирования досталось данное ТЗ. В предположении о нормальном законе распределения результатов тестирования, наибольшая точность определения индекса дискриминации ожидается при объеме каждой из двух крайних групп в 27 % от общего числа испытуемых;  $N_{n_{\min}}$  – число испытуемых в нижней крайней группе, успешно справившихся с данным ТЗ;  $N_{n_{\max}}$  – число студентов в верхней крайней группе, успешно справившихся с данным ТЗ;

Значения коэффициента дискриминативности ТЗ рассматриваемого модуля представлены на рис. 4. Если считать приемлемым значения указанного показателя величину 0.4 (тонкая линия), то качество нескольких десятков ТЗ при подготовке к входному тестированию в 2022 году пришлось улучшить.

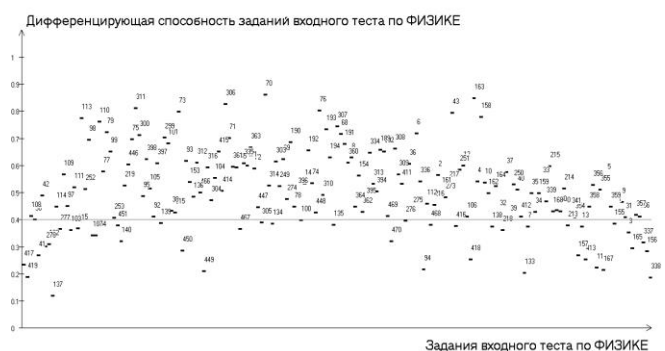


Рис. 4. Значения коэффициента дискриминативности ТЗ модуля «Механика» (27-процентные крайние группы)

Как и ожидалось, основной (но не единственной) причиной низкой дифференцирующей способности ТЗ стал слишком низкий или слишком высокий уровень их сложности. Проблема состоит в том, что в момент составления тестового задания (априорно) невозможно оценить будущую (апостериорную) оценку показателей его качества, и по результатам апробации у многих (если не у большинства) заданий выявляется неполное соответствие указанным критериям. Просто заменить такие задания на вновь составленные невозможно, поскольку многие (возможно – большинство) вновь предложенных также не пройдут экспериментальную проверку. Единственный конструктивный подход состоит в коррекции заданий на основе анализа результатов тестирования многочисленных студентов. Например, если некоторое задание по «Математике» из категории «простых» вызывает затруднение у необъяснимо большого числа студентов, его необходимо передать специалистам для ручного анализа. Предположим, что это задание подразумевает словесный ответ «неопределенный интеграл». Тогда у всех

студентов, давших ответ «неопределённый интеграл», он не будет зачтен.

Решение о коррекции баз тестовых заданий – а именно о размещении в ней второго, альтернативного варианта правильного ответа, не может быть принято без участия человека – эксперта, что наглядно иллюстрирует справедливость отнесения систем индивидуализации образования к «физической сфере».

И только после того, как качество баз тестовых заданий достигнет необходимого уровня, а результатам входного компьютерного тестирования начинают доверять все заинтересованные участники учебного процесса, может идти речь об изменении образовательных технологий, применяемых в ПНИПУ, в соответствии с реальным освоением первокурсниками тех или иных элементов школьной программы. На рис. 5 показана доля правильных ответов на ТЗ, относящиеся к различным темам раздела «Динамика поступательного движения».

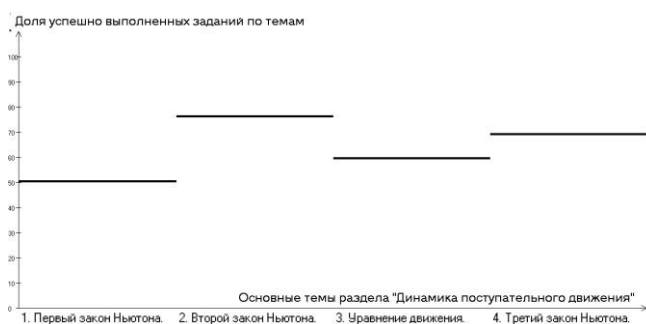


Рис. 5. Результаты входного компьютерного тестирования студентов ПНИПУ в сентябре 2022 года

Неожиданно низкие результаты по теме «Первый закон Ньютона» методисты ПНИПУ склонны объяснять неоправданно малым временем, выделяемым школьной программой этой, ошибочно считающейся очень простой в освоении, теме. Очевидно, что подобный вывод будет объективным лишь при условии полной уверенности в том, что относящиеся к данной теме ТЗ не являются чрезмерно сложными. А это, в свою очередь, возможно исключительно в результате анализа статистически значимого числа ответов первокурсников на задания теста. Поскольку же база тестовых заданий должна обладать свойством репрезентативности (хотя бы для того, чтобы исключить частое повторение одного и того же задания у студентов, тестирующихся за соседними

компьютерами), суммарный объем обрабатываемой информации с полным правом позволяет относить предметную область к категории Big Data.

Еще большую размерность имеет задача формирования списков студентов, показавших недопустимо низкие результаты освоения школьной программы. Для этих студентов организуются дополнительные занятия с целью «дотянуть» их до необходимого уровня понимания изучаемого в ВУЗе материала. Однако не целесообразно проводить такие занятия со всеми недостаточно подготовленными студентами по всем разделам и темам школьной программы. Каждый отдельный студент должен посетить лишь те из них, которые связаны с конкретными «пробелами», показанными им на входном тестировании, что требует составления индивидуального расписания, причем согласованного с расписанием учебных занятий каждой специализации или профиля подготовки в ПНИПУ. Индивидуализация корректирующей подготовки позволяет значительно сократить затраты на оплату труда преподавателей, а также сэкономить аудиторный фонд.

#### IV. ЗАКЛЮЧЕНИЕ

Рассмотренный подход индивидуализации образовательных траекторий существенно повышает планку требований к объективности процедуры компьютерного тестирования. А это, в свою очередь, подразумевает необходимость близкого к идеальному уровню качества баз тестовых заданий как главного компонента в системе педагогических измерений. И только после детальной доработки качества баз тестовых заданий становится возможным и экономически оправданным использование результатов входного тестирования для формирования индивидуального подхода к учебной работе каждого студента.

#### СПИСОК ЛИТЕРАТУРЫ

- [1] Lynch С.А. Big data: How do your data grow? // Nature. 2008. Vol. 455. No 7209. P. 28–41.
- [2] Новиков Д.А. Большие данные: от Браге – к Ньютону // Проблемы управления. 2013 г. №6. С.15-23.
- [3] Большие данные в образовании: доказательное развитие образования: Сб. науч. статей II Международной конференции, Москва, 15 октября 2021 г., с. 343.
- [4] Фиофанова О.А. Анализ больших данных в сфере образования: методология и технологии: монография М.: Издательский дом «Дело» РАНХиГС, 2020 г., с. 200.