

ИЗВЛЕЧЬ СМЫСЛ

проблемы анализа данных в образовании

Алексей Семенов, Ольга Фиофанова, Ольга Бабченко,
Илья Бронштейн, Нияз Габдрахманов

АННОТАЦИЯ В октябре 2021 года состоялась II Международная научно-практическая конференция «Большие данные в образовании: доказательное развитие образования». На мероприятии прошел круглый стол по проблематике архитектуры данных и аналитики данных в образовании. Предлагаем вашему вниманию фрагменты докладов некоторых участников.

КЛЮЧЕВЫЕ СЛОВА Большие данные в образовании, архитектура данных, интерпретация данных.

DOI 10.22394/2078-838X.2021-3-60-64

УЧАСТНИКИ КРУГЛОГО СТОЛА

Алексей Львович СЕМЕНОВ

д. физ.-мат. н., руководитель программы РФФИ,
академик РАН, академик РАО

Ольга Александровна ФИОФАНОВА

д. пед. н., руководитель проекта РФФИ

Ольга Ивановна БАБЧЕНКО

Заместитель начальника
Департамента образования Екатеринбурга

Илья Михайлович БРОНШТЕЙН

Министр образования Московской области

Нияз Камилевич ГАБДРАХМАНОВ

к. г. н., эксперт, научный сотрудник Института
образования НИУ Высшая школа экономики

Показатели качества

Алексей Семенов. За последние годы московское образование сделало огромный шаг вперед в сторону цифровизации. В частности, «Московская электронная школа» стала общим цифровым пространством для всех учителей, учащихся и родителей нашего города.

В течение десятилетий Российский фонд фундаментальных исследований был источником грантового финансирования для ученых, занимающихся исследованиями в области точных, естественных и гуманитарных наук. То, что мы в конце 2019 года, непосредственно перед тем, как разразилась пандемия, запустили программу фундаментальных исследований цифровой трансформации российской школы, оказалось предвидением. Сейчас мы находимся в середине этого большого проекта, в котором участвует более 60 коллективов из многих регионов России.

О чем именно идет речь, когда мы говорим «большие данные»? Сегодня технические средства позволяют собирать огромную информацию о ходе образовательного процесса. Например, до пандемии учителя литературы говорили: «Зачем писать сочинения на компьютере?» Но уже год назад они стали говорить: «Конечно, дети сдают сочинения в цифровом виде!». Благодаря переходу в цифровой формат учитель получает возможность содержательно комментировать тексты. Учащийся может учитывать эти комментарии.

Собирая данные о каждой образовательной организации, о каждом учителе и учащемся, мы получаем объективную возможность доказательно оценивать работу человека в сравнении с ним самим, работу школы в сравнении с ней самой.

Как из больших данных извлечь показатели качества? Это очень важная задача. На самом деле, я думаю, что мы приближаемся к ее решению. Если мы сумеем это сделать, то, например, может оказаться, что лучшая школа Москва – та, в которой постепенно двоечников сменили четверочники.

Хотелось бы также подчеркнуть, что мы получаем большую помощь, в частности, от Департамента образования и науки города Москвы и от учителей в создании учебника «Цифровой мир». Большое спасибо.

Уровни анализа данных

Ольга Фиофанова. Рассуждая об извлечении смысла из аналитики данных, мы должны понимать, что смысл зависит от двух факторов: технологического (инфраструктура, архитектура данных в цифровых средах, стандарты анализа данных) и антропологического (обращение человека к данным для обоснования решений, знание методов анализа данных, понимание практического применения data-анализа).

Извлечение смысла из анализа данных – это человеческий фактор. Привычка обращаться к данным преобразуется в соответствующий способ мышления, основанный на построении информационных моделей, выявлении связи и структуры данных и оперировании ими.

А для педагога это еще и сверхчеловеческая задача: создать на основе аналитики данных новые знаниевые системы – образовательные программы для развития мышления, data-грамотности, data-компетенций.

Таким образом, цикл управления данными связывается с циклом управления знаниями.

И в этом – организационном – смысле – было бы правильным, полагаю, строить корпоративные системы обучения с использованием аналитики данных региональных систем образования. Такое организационное решение уже осуществлено, например, в Москве и Подмосковье: созданы, например, Корпоративный университет Правительства Московской области и Корпоративный университет Московского образования, ставший

одним из победителей Всероссийского конкурса кейсов по анализу данных и доказательному развитию образования.

Это история о том, как аналитика данных в форме датасетов интегрируется в корпоративную систему обучения и развития компетенций педагогов и руководителей образования.

Обсуждая проблематику анализа данных и смыслообразования на основании такой аналитики, нужно понимать различие в уровнях управления и оперирования данными для развития образования.

Разные уровни управления – разные задачи анализа данных – разные смыслы для доказательного развития национальных, региональных систем образования, образовательных институтов и человеческого потенциала.

1. На уровне государственного управления объектом анализа являются данные по индикаторам национальных проектов и государственных программ развития образования, показатели мониторингов системы образования, данные Национальной системы оценки качества образования.

Технологическая инфраструктура данных этого уровня анализа:

- а) Национальная система управления данными – ЕИП НСУД / наборы данных
- б) Портал госпрограмм – сводная аналитика
- в) OECD / Open Data и другие.

Основной проблемой этого уровня управления на основании данных остаются вопросы и вытекающие из них смыслы:

- как институционализируется «человеческое измерение» (гуманитарная аналитика) в системе управления на основании данных?
- как изменяются методики и индикаторы мониторингов образования, чтобы наиболее корректно отражать задачу «равные возможности для каждого»?
- как встраиваются инструменты Data-Anthro в цифровые сервисы аналитики данных на отраслевом/межотраслевом, территориальном (региональном), государственном уровне?

2. На уровне организационного управления объектом анализа являются данные, отражающие развитие и управление образовательной организацией или управление образовательной платформой (например, РЭШ, МЭШ, ГИС СЦОС)

Технологическая инфраструктура данных этого уровня анализа: данные программ развития

¹ Конференция прошла при поддержке Российского фонда фундаментальных исследований (проект 19-29-14016, руководитель О. А. Фиофанова).

образовательных организаций, данные мониторинга системы образования данные по организациям, данные о развитии образовательной организации, ее кадровом потенциале, достижениях ее обучающихся, например в ЛК руководителя по мониторингу «Надежная школа».

Основной проблемой этого уровня управления остаются вопросы и вытекающие из них смыслы об институциональных (или платформенных) условиях развития человеческого потенциала, об институциональных данных качества образования, о связи аналитики данных с качеством преподавания и учения.

3. На уровне педагогической аналитики данных и педагогической организации индивидуальных образовательных траекторий, индивидуальных учебных планов объектом анализа являются данные, отражающие развитие человека, данные выборов, осуществляемых человеком среди образовательных программ, способов решения учебных и проектных задач, учебных затруднений человека, его мотивов и предпрофессиональных, жизненных устремлений.

Технологическая инфраструктура данных этого уровня анализа: архитектура данных в LMS университета / школы / образовательной организации.

Основной проблемой этого уровня управления остаются вопросы и вытекающие из них смыслы о стандарте цифровой образовательной среды и связанных с этим вопросом неравных возможностей обучения в средах технологически развитых с возможностью построения индивидуальных образовательных траекторий, рефлексией школьником / студентом собственных данных; по сравнению с цифровыми образовательными средами технологически неразвитыми, где таких возможностей нет). А также принципы аналитики данных в условиях смены модели образования и переходу к модели «Образование 4:0» («Дидактика 4:0»).

Для этого важно различать:

- Образование индустриальной эры — это предметно организованный учебный материал, контроль и анализ данных по освоенному объему знаний и репродуктивным навыкам — с соответствующей аналитикой данных.

- Образование постиндустриальной эры — это конвергентно организованный образовательный контент (содержание образования), деятельностная оценка и анализ данных по развитию продуктивной деятельности, анализ данных личных выборов в решении проектных задач. Так как кроме

традиционных культурных норм важно развивать у поколения next собственную культурную норму управления собственным образованием, развить на протяжении всей жизни.

Глубоко понимая смыслы вышеназванного, мы организовали разнообразные формы поддержки культуры data-анализа:

1) Международная конференция «Анализ данных в образовании: доказательное развитие образования»;

2) Всероссийский конкурс кейсов по анализу данных и доказательному развитию образования;

3) Образовательную программу высшего образования «Управление образованием на основании данных. Проект получил финансовую поддержку Фонда Владимира Потанина и стал победителем конкурса на предоставление грантов преподавателям магистратуры 2020/2021 благотворительной программы «Стипендиальная программа Владимира Потанина» благотворительного фонда Владимира Потанина.

В настоящее время эта новая программа магистратуры готовится к аккредитации. Мы будем рады увидеть в числе магистрантов этой программы читателей журнала «Образовательная политика».

Организационные формы поддержки мы развивали также на основе анализа данных, но в данном случае — Data-Science — научных данных. Результаты проекта 19-29-14016 фундаментального научного исследования «Методология анализа данных и ее интеграция в программы профессиональной подготовки педагогов и руководителей образования», поддержанного грантом Российского фонда фундаментальных исследований <https://rfi.1sept.ru/person/450>.

Данные и смысл

Ольга Бабченко. Екатеринбург — один из крупнейших городов России. Система образования мегаполиса в процессе работы обрастает огромным количеством данных и мониторинговых систем. Но, как известно, данные — это еще не все. Важно извлекать из них какой-то смысл. Вопрос об оптимизации работы со значительным объемом накопленных данных — один из самых актуальных в управленческой деятельности на муниципальном уровне.

Большие данные позволяют решать важные задачи, например, оценивать особенности функционирования образовательного комплекса,

оперативно реагировать на изменения и отклонения в результатах и определять перспективы развития, повышать управляемость системы. Одной из значимых характеристик системы на современном этапе является то, что образование уже давно вышло за рамки образовательной организации. Это происходит через интеграцию систем общего и дополнительно образования, через введение в эксплуатацию различных инновационных структур, таких как «Кванториумы», IT-кубы, через реализацию сетевых образовательных программ. В Екатеринбурге эти процессы развиваются очень активно.

Все наши потоки данных так или иначе ориентируются на нашу внутреннюю систему. Выходить за рамки образовательной организации нам пока очень сложно. Мы еще не отработали конкретные механизмы.

В системе образования Екатеринбурга работает несколько автоматизированных информационных систем, самой крупной из них является так называемая АИС «Образование», которая эксплуатируется и развивается более 15 лет. Мы думаем, что мы уже создали мощного «монстра», который содержит огромное количество модулей. Из этой системы мы можем взять статистические данные, которые касаются сети образовательного комплекса и контингента, которые говорят о результатах муниципальной оценки системы качества образования, о результатах участия школьников во всероссийских олимпиадах, о динамике и результатах физической подготовки обучающихся. В общем, по разным направлениям.

Объем данных огромен, но он нас не удовлетворяет, поскольку во внешнее пространство системы мы выходим по минимальным показателям.

Какие риски и системные проблемы муниципального уровня для обработки больших данных мы видим? Проблемой является недостаток финансовых ресурсов для модернизации существующих систем и их интеграции с федеральными и региональными платформами и автоматизированными системами для обеспечения больших объемов дискового хранилища. Отсутствует достаточное количество квалифицированных специалистов и экспертов для анализа статистических данных и получения выводов. Сложность интерпретации данных на муниципальном

ДАННЫЕ — ЭТО ЕЩЕ НЕ ВСЕ. ВАЖНО ИЗВЛЕКАТЬ ИЗ НИХ СМЫСЛ

уровне возникает из-за многообразия особенностей функционирования учреждений.

Что мы видим для себя в качестве перспективных направлений и задач использования больших объемов данных на муниципальном уровне? Ну, во-первых, это межведомственная система учета контингента. Нам это интерес-

но, потому что речь идет об учете демографических тенденций и миграционных процессов. Кроме того, у нас количество учеников десятых классов сокращается. Мне кажется, межведомственная система учета контингента помогла бы нам обосновывать необходимость создания классов на базе той или иной образовательной организации.

Мы очень надеемся на интеграцию внутренних и внешних потоков данных. Например, актуальной для нас является задача интеграции муниципальной и федеральной систем оценки качества.

Необходимо использовать автоматизированные информационные системы, чтобы оперативно решать вопросы, возникающие у населения. Сейчас мы начали отрабатывать такой опыт. Например, при изменении законодательства в части регулирования перевода детей из одного дошкольного учреждения в другое заявитель должен в обязательном порядке согласовывать все эти моменты с учредителем.

Проблема интерпретации

Илья Бронштейн. Существующие системы, будь то электронный дневник или система записи в школу, часто напоминают монстров, потому что обрастают гигантскими модулями, каждый из которых генерирует какое-то количество информации. Это приводит к перегрузке информационной системы, потому что по прошествии нескольких лет никто уже не помнит, какие смыслы закладывались в тот или иной модуль.

Мы научились собирать и генерировать данные, мы научились видеть между ними связи. Но какие выводы мы можем делать на основе этих данных? Весной прошлого года мы, как и вся страна, были вынуждены перейти на дистанционное обучение. Мы поняли, что цифровая интерпретация результатов выполненных учениками тестов часто отличается от той

интерпретации, которую учитель дает в классе. Можем ли мы сказать, что автоматическая оценка более достоверна, чем та, которую поставил учитель? Уверен, что нет. Уверен, что мнение учителя должно главенствовать.

Но как тогда сопоставить мнение учителя с результатами, которые получил ребенок? Я не знаю действенных инструментов, позволяющих в системе образования анализировать большие данные и делать достаточно достоверные выводы. Хочу подчеркнуть, что для Московской области такие решения были бы очень важны. У нас сегодня почти миллион школьников, у нас сложная социальная и территориальная структура. У нас, например, есть школы, расположенные в труднодоступных населенных пунктах, есть большие городские школы, сельские школы и т. д.

И понятно, что структура данных, необходимых для принятия тех или иных управленческих решений, тоже будет разной. И поэтому для нас очень важно найти такие инструменты. Мы в поиске. Московская область готова обсуждать решения, готова к сотрудничеству.

Механизм деперсонализации

Нияз Габдрахманов. Мы прекрасно знаем практики работы с большими данными университетов и образовательных учреждений другого уровня. Эти организации генерируют колоссальное количество данных. С точки зрения ресурсной составляющей мы эту проблему фактически решили. Проблема заключается в другом. К сожалению, эти данные никак не используются, на их основе не принимаются управленческие решения.

Несколько лет подряд мы проводим исследования, посвященные анализу социального самочувствия студентов, и оперируем данными социальной сети «ВКонтакте».

Мы собираем цифровые следы в рамках консорциума. Вузы становятся инициаторами работы с большими данными. На мой взгляд, это важный сдвиг.

Вместе с тем есть очень важные ограничения с точки зрения использования этих данных. Все-таки данные, которые генерируются образовательными учреждениями, носят некий индивидуальный характер. Нужно искать механизм, который позволит их деперсонализировать.

Вторая проблема: как шить разные форматы таблиц и данных? Ведь зачастую в образовательных учреждениях данные генерируются очень разнородные по структуре.

Еще одна проблема заключается в том, что данные, которые генерируются внутри образовательного учреждения, как правило, находятся в разных структурных подразделениях. Данные по школам хранятся в Министерстве просвещения, данные по вузам – в Министерстве науки и высшего образования. Межведомственное взаимодействие по обмену данными налажено пока слабо. При этом, когда мы говорим об образовательной успешности, важно наблюдать поступление человека в школу, его обучение в школе, в вузе.

Поэтому я хотел бы акцентировать внимание на этих трех основных компонентах: работа с персональными данными, структура данных и, конечно же, межведомственные взаимодействия как внутри образовательных учреждений, так и между министерствами ведомств.

EXTRACT MEANING problems of data analysis in education

Alexei L. SEMENOV.

Prof., Head of the RFBR program, Academician of the Russian Academy of Sciences, Academician of the Russian Academy of Education.

Olga A. FIOFANOVA.

Doctor of Pedagogical Sciences, Head of the RFBR project.

Olga I. BABCHENKO.

Deputy Head of the Yekaterinburg Department of Education.

Илья М. BRONSTEIN.

Minister of Education of the Moscow region.

Niyaz K. GABDRAKHMANOV.

PhD, expert, researcher at the Institute of Education of the Higher School of Economics.

Abstract

In October 2021, the II International Scientific and Practical Conference "Big Data in Education: Evidence-based development of education" was held. A round table on the problems of data architecture and data analytics in education was held at the event. We bring to your attention fragments of reports of some participants.

Keywords Big data in education, data architecture, data interpretation.

