

ДАТА-ГРАМОТНОСТЬ

Модная тема или необходимость?

Павел Глухов, Андрей Дерябин, Александр Попов

АННОТАЦИЯ В статье рассматривается понятие дата-грамотности. Исследуется содержание и состав такого вида грамотности, обосновывается ее актуальность как образовательного результата, сообразного трендам цифровизации. Разделяются подходы к углубленному и общему изучению науки о данных. Дается описание различных типов задач, поставленных на разном учебном материале и направленных на развитие дата-грамотности у учащихся. Рассматриваются способы разворачивания программ, направленных на освоение учащимися науки о данных без необходимости оформления ее в отдельную дисциплину.

КЛЮЧЕВЫЕ СЛОВА Наука о данных, машинное обучение, образовательные программы, дата-грамотность, анализ данных, data science, общее образование, профориентация, дополнительное образование, информатика.

DOI 10.22394/2078-838X-2020-3-74-82

Датафицированное общество

Сегодня мы живем в мире, где существенная часть повседневности включает в себя разные массивы данных. Благодаря развитию технологий машинного обучения, в основе современной глобальной экономики лежит процесс сбора и обработки данных. Помимо прочего собираются и обрабатываются персональные данные, которые позволяют сделать сферу услуг более индивидуализированной. Все мы сталкивались с таргетированной рекламой, которая появляется в социальных сетях при просмотре ленты новостей. Такая реклама формируется на основании ряда критериев (пол, возраст, уровень потребления, место жительства, интересы), которые устанавливаются на основе персональных данных пользователя в результате мониторинга его цифровых следов (что он «лайкает», какие запросы делает в поисковиках, на какие паблики в социальных сетях подписан и т. д.). Уже давно не является секретом то, что персональная информация собирается и обрабатывается корпорациями, а таргетированная реклама — это лишь один из наиболее очевидных примеров, который демонстрирует, что данные, не только персональные, являются «новым золотом XXI века».

Для многих подобные обстоятельства рисуют картину антиутопии в жанре киберпанка, где проворные корпорации все про нас знают и устраивают ежедневную персональную

слежку при помощи камер за каждым сделанным шагом, а при помощи платежных систем — за каждой транзакцией. Это не так далеко от правды, если основываться на опыте Китая, где действует система социального рейтингования граждан, основанного на данных, собираемых компаниями Alibaba и Tencent. Однако наука о данных (data science) установила новую планку качества в сфере прогнозирования по разным направлениям.

«ДУМАТЬ ДАННЫМИ» — ЭТО СВОЕГО РОДА МЕТАНАВЫК

При помощи анализа больших данных выявляется множество закономерностей, имеющих как общекультурную, так и финансовую ценность. Например, Альберт-Ласло Барабаши на основе больших данных предсказал взлеты репутации художников в зависимости от траектории проведения их выставок в сложной и запутанной сети галерей и музеев. Это также позволило спрогнозировать стоимость картин. Да, прогнозировать стоимость картины умели при помощи экспертов, без использования науки о данных. Но разница заключается в том, что сейчас практически каждый может сделать это при помощи открытого и апробированного инструментария. Можно сказать, что



**Павел Павлович
ГЛУХОВ**

Научный сотрудник научно-исследовательского сектора «Открытое образование» научно-исследовательского центра социализации и персонализации образования детей Федерального института развития образования (ФИРО) РАНХиГС, эксперт лаборатории компетентностных практик образования МГПУ (125319, РФ, Москва, ул. Черняховского, д. 9, стр. 1).
E-mail: gluhovpav.pav@gmail.com

**Андрей Александрович
ДЕРЯБИН**

MSc Social Psychology, научный сотрудник научно-исследовательского центра социализации и персонализации образования детей ФИРО РАНХиГС (125319, РФ, Москва, ул. Черняховского, д. 9, стр. 1).
E-mail: deryabin-aa@ranepa.ru



**Александр Анатольевич
ПОПОВ**

Д. филос. н., доцент, зав. научно-исследовательским сектором «Открытое образование» научно-исследовательского центра социализации и персонализации образования детей ФИРО РАНХиГС. Зав. лабораторией компетентностных практик образования Института системных проектов МГПУ. Профессор кафедры социологии и массовых коммуникаций гуманитарного факультета Новосибирского государственного технического университета (125319, РФ, Москва, ул. Черняховского, д. 9, стр. 1). E-mail: aktor@mail.ru

Барабаши десакрализировал это знание. Подобные закономерности он также выявил в сфере книгоиздания, где анализировал траектории успеха авторов бестселлеров и определял вероятность воспроизводства их популярности. Также группы ученых при помощи науки о данных выявили факторы успеха суперзвезд в мировом Dj-сообществе (Janosov & Muscioto et al., 2019), факторы успеха в индустрии стартапов (Bonaventura & Ciotti et al., 2019) и т. д. Таким образом, наука о данных дает конкретные и статистически обоснованные «формулы успеха» с попутным выстраиванием картины функционирования интересующей индустрии, в отличие от популярных тренингов личностного роста.

Уже сформирован большой спрос на специалистов, обладающих компетенциями в сфере data science, но стоит понимать, что данные — это не просто объект цифровой экономики, а неотъемлемый элемент современной картины мира. Данных становится все больше, они могут рассказать нам о нас самих¹ и объяснить многие актуальные феномены².

Дата-грамотность

Экспансия больших данных во все сферы жизни заставляет задуматься о необходимости популяризации данной темы на массовом уровне. Необходимо ли изучать науку о данных наравне с физикой или обществознанием?

Прежде всего стоит понимать, что наука о данных, как и любая область знания, имеет разные уровни специализации и сложности. Сегодня мы можем встретить большое количество предложений на образовательном рынке, которые направлены на обучение программированию и анализу данных. Широко распространены онлайн-курсы, которые пытаются реагировать на кадровый заказ. Примеры курсов на платформах Coursera и edX:

<https://www.coursera.org/browse/data-science>,

<https://www.edx.org/course/subject/data-science>.

¹ Например, выявлены новые закономерности, влияющие на прочность социальных связей (Navarro & Miritello et al., 2017).

² Одним из самых ярких примеров является критика концепции «долгого мира» на материале больших данных (Cirillo & Taleb, 2016a).

Также мы можем увидеть большое количество программ профессиональной подготовки в области data science на уровне бакалавриата и магистратуры (De Veaux, 2017; Baumer, 2015; Anderson, 2014). Однако данные примеры можно отнести к программам углубленного изучения, которые интересны специалистам, но могут быть избыточны для широких масс. Такие программы сфокусированы на изучении технической стороны анализа данных и уделяют меньше внимания общему влиянию данных на жизнь общества и человека, что также отмечается экспертами (Heinemann et al., 2018).

Инструменты data science являются более многогранными и доступными, чем может показаться на первый взгляд. В академическом сообществе обсуждается понятие дата-грамотности (data literacy) как способности общего характера, применимой в рамках общественной и этической проблематики, позволяющей человеку принимать рациональные решения и вырабатывать подкрепленную позицию в современном цифровом мире (Schield, 2018). Дата-грамотность представляется как комплексная характеристика современного информированного гражданина и может рассматриваться с трех базовых позиций:

- как позволяющая человеку оценивать последствия использования и предоставления данных о себе и своей жизни при работе с платежными системами, онлайн-сервисами, использовании телефона, разнообразного программного обеспечения и т. д.;
- как позволяющая читать и интерпретировать статистику, которую мы можем увидеть

в общественных, экономических и политических процессах, повестках, новостных сводках;

- как позволяющая добывать и обрабатывать данные, необходимые для более объективного принятия жизненных решений, которые могут быть связаны с выбором рабочего места, города проживания или с прогнозированием финансовой ситуации собственной семьи.

Дата-грамотность предполагает владение ин-

струментами количественного и качественного анализа данных, но гораздо более важным умением здесь выступает применение подобных инструментов в определенном контексте (Gibson & Mourad, 2018). Мы часто обнаруживаем разные статистические данные, которые могут иметь определяющее

значение для нашей повседневной жизни — от курсов валют и информации о распространении COVID-19 до влияния экологии или даже ожирения на смертность в регионах. Для того чтобы иметь возможность критически ознакомиться с подобными статистическими обобщениями, необходимо иметь опыт или знание о том, как происходит работа с подобными данными. Pangrazio и Sefton-Green (2019) утверждают, что к наборам данных следует относиться осторожно, так как они могут претерпевать слишком много циклов обработки и использоваться разными агентами в разных контекстах, с разными целями и мотивами. В этом смысле одна из важнейших задач формирования дата-грамотности — это развитие умения реконструировать преподносимую информацию

С БОЛЬШИМ УСПЕХОМ МЫ МОЖЕМ ПРИМЕНЯТЬ АНАЛИЗ ДАННЫХ В РАМКАХ ЛИТЕРАТУРЫ

Термин Big Data — «большие данные» — был впервые озвучен в 1998 году в публичном докладе Джона Мэши, ведущего инженера компании Silicon Graphics. С тех пор объемы данных, производимых людьми, только возрастают. Возрастает и потребность в умении работать с ними. При этом, согласно исследованию The Human Impact of Data Literacy, проведенному компаниями Qlik и Accenture в 2020 году, только 21 % молодежи от 16 до 24 лет обладает грамотностью по работе с данными.



до уровня используемых в ней данных и оценивать контекст использования. Например, многие столкнулись с трудностями при чтении ежедневных статистик о распространении COVID-19. В сводах по распространению пандемии мы часто видели прогнозы, содержащие чуть ли не дату завершения пандемии. Однако без понимания того, на каких данных основаны подобные выкладки и какие факторы там не учтены, также сложно дать адекватную оценку такой информации.

Мы должны рассматривать дисциплину «Наука о данных» в контексте решения актуальных для ученика задач и проблем. Возьмем типичную для большинства подростков жизненную ситуацию: поступление в учебное заведение. Можно решать подобную задачу множеством способов. Один из них — это спросить мнение учителей, родителей, посмотреть информацию на сайтах и в буклетах. При таком подходе мы часто упираемся в критерий репутации вуза, и, следовательно, чем выше статус вуза, тем мы более склонны желать поступить в него. Но будут ли приведенные нами источники данных действительно что-то говорить о репутации заведения, и как ее можно оценить самостоятельно, не прибегая к дорогостоящим опросам?

Например, при помощи выгрузки данных, находящихся в открытых источниках, и создании алгоритма, автоматизирующего этот процесс (парсинг данных), мы можем исследовать публичную репрезентацию конкретных вузов, не затрачивая на это большого количества времени. Установить, о каком вузе и с какой частотой отзываются позитивно, а о каком негативно. Созданный алгоритм обрабатывает потоки открытых данных по интересующим нас параметрам. Например, несколько миллионов публичных сообщений в социальной сети. При этом мы задаем параметры, которые интересуют именно нас (например, пол, возраст или место обучения). Потоки данных собираются в рамках интересующего нас периода. Например, нас может интересовать период весна — лето, так как это активный период среди абитуриентов. Мы находим в данных именно те, которые представляют для нас ценность по заданным в алгоритме параметрам. Например, мы хотим понять, как часто люди отзываются позитивно о вузе, и при анализе данных применяем критерий частоты употребления слов «я люблю», «мне нравится»,

«обожаю» и т. д. в связке с названием вуза и получаем частоту употребления.

Помимо сообщений мы можем анализировать иные единицы, говорящие о публичной репрезентации. Например, мы можем исследовать чекины — отметки о посещении конкретной локации, которыми пользователи сопровождают свои посты в социальных сетях. Безусловно, чекины не отражают ситуацию реального трафика посещаемости места. Но позволяют проанализировать содержимое самого поста, где мы можем судить об удовлетворенности пользователя посещением. Также чекины распространяют информацию о месте среди других пользователей (друзей пользователя), так что их значимость потенциально важнее, чем данные о трафике.

Как формировать дата-грамотность?

Finzer считает, что «думать данными» — это своего рода метанавык, и его развитие в школе должно происходить через применение методов науки о данных в рамках существующих предметов, а не через введение специальной дисциплины: «Два наблюдения показывают, что наука о данных не должна стремиться стать новой предметной областью в школьной программе — в ней и так уже больше предметов, чем это представляется комфортным, и — обучение работе с данными имеет наибольший смысл в релевантном [предметном] контексте» (Finzer, 2013). Как отдельная дисциплина наука о данных уместна в профессиональной или предпрофессиональной подготовке. Науку о данных следует понимать как средство, позволяющее познавать другие дисциплины. Очевидно, что сейчас в рамках школьной подготовки реализация метадисциплинарного развития дата-грамотности у учеников представляется маловероятной, так как требует переосмысления содержания образовательных программ большинства дисциплин. Но если бы нам дали такую возможность, то как это могло бы выглядеть?

Посредством case study

Сейчас данный метод представляется наиболее вероятным, так как носит фрагментарный характер. Можно взять один кейс, внедрить его

в качестве элемента программы там, где он будет более уместен тематически. В качестве возможного и многогранного примера здесь может выступить дискуссия вокруг концепции «долгого мира»³ авторства С. Пинкера. В данной концепции на основании массивов статистических данных выстроено доказательство того, что в XXI в. фиксируется глобальный спад насилия и нам не следует ожидать глобальных военных конфликтов. Данная концепция достаточно популярна как раз за счет обширных статистических выкладок, которые создают ощущение достоверности. Она критиковалась множеством журналистов и ученых, однако встретила серьезное сопротивление только в лице Нассима Талеба в соавторстве с Паскуале Сирилло (Cirillo & Taleb, 2016b), которые противопоставили другую математическую модель с более широким и глубоким набором данных, опровергая концепцию Пинкера.

Это видится как очень важная на сегодняшний день дискуссия, которая затрагивает множество предметов⁴. Погружаясь в разбор данного кейса, учащиеся прежде всего могут разделить на две «команды»: за Пинкера или за Талеба. Основание для такого разделения — это симпатия не только к автору (хотя подобное не исключается), но и к критериям, заложенным в одну из двух математических моделей. Реконструируя информацию, на которой построены доказательства авторов, мы получаем возможность охватить с учащимися разные эпохи и через статистические данные графиков последовательно изучить интересующие нас исторические события: в частности, войны, их масштабы и законы воспроизводства в нашей цивилизации. При этом этическое содержание концепций Пинкера и Талеба может касаться общественнознания, так как они пытаются объяснить гуманитарный феномен насилия. Подтверждать ценность кейса с позиции точных наук не приходится, так как речь идет о математических моделях. В сущности, попытавшись осуществить реконструкцию одной из моделей в рамках данного кейса, можно

освоить комплексное представление о том, что такое большие данные, как они формируются и применяются.

Посредством попутных задач

Если реализация кейса по разбору концепции «долгого мира» представляется как достаточно комплексная и последовательная работа, то можно прибегать к более простым и еще более точечным приемам при формировании дата-грамотности. Например, ставя перед учениками задачи, которые позволяют «поиграться» с учебным материалом. С большим успехом мы можем применять анализ данных в рамках литературы. Так как текст также подвергается количественной обработке, мы можем придумывать и применять модели разной степени сложности по интересующим нас критериям, в том числе для удовлетворения обычного любопытства. Мы можем взять произведения разных авторов и сравнить их между собой на предмет упоминаемости какого-либо комплекса слов или их оттенков. Таким образом мы можем подтвердить или опровергнуть гипотезу о том, какой автор писал более депрессивные тексты, а какой более жизнерадостные. Сопоставив результаты анализа по разным авторам, получаем возможность сформировать рейтинг депрессивности авторов.

Следует заметить, что методы количественного анализа, которые применяются в науке о данных, не во всех случаях обязательно нужно применять именно к большим массивам данных.

Посредством образовательных задач

Учащиеся могут получать или разрабатывать самостоятельно задачи, которые связаны с их жизненным самоопределением или познавательным интересом. Дальнейшее решение таких задач можно осуществлять в разных формах самостоятельной работы (например, научно-исследовательской или проектной работы). Учитель здесь выступает как консультант, удерживающий свою предметность. В рамках решения такой задачи учащийся должен пройти четыре фазы разработки: (1) понятийная (формирование базового понятия в рамках задачи, выделение различных фокусов его рассмотрения), (2) аналитическая (выделение объектов анализа;

3 «The Better Angels of Our Nature»: https://en.wikipedia.org/wiki/The_Better_Angels_of_Our_Nature.

4 С учетом того, что в рамках данной дискуссии при помощи математических методов впервые было доказано, что используемые исторические данные о военных потерях в разы преувеличены.

количественный, статистический анализ ситуации согласно фокусам рассмотрения темы), (3) фаза моделирования (разработка инфографики, прогнозирование тенденций развития согласно фокусам рассмотрения темы на ближайшие годы), (4) рефлексивная фаза (написание тематического эссе с интерпретацией результатов анализа, рассмотрением их прикладного значения, выступление с докладом).

Задачи такого рода могут быть поставлены в рамках совершенно разных предметных областей. Например, основываясь на статистических данных, мы можем исследовать культурный уровень нашего города или региона через постановку вопроса о том, в какой степени успешность городского/регионального развития определяется не только экономическими, но и культурными факторами? Как можно измерить такую нематериальную вещь, как культура? Учащиеся знакомятся с данными, выполняя простейшие аналитические действия: какой регион (по какому показателю) имеет наивысшее/низшее значение? Почему? Что мы знаем об этом регионе? Между какими культурными, экономическими, социальными, популярными

УЧАЩИЕСЯ МОГУТ ПОЛУЧАТЬ ИЛИ РАЗРАБАТЫВАТЬ САМОСТОЯТЕЛЬНО ЗАДАЧИ, КОТОРЫЕ СВЯЗАНЫ С ИХ ЖИЗНЕННЫМ САМООПРЕДЕЛЕНИЕМ ИЛИ ПОЗНАВАТЕЛЬНЫМ ИНТЕРЕСОМ

ляционными и прочими признаками может существовать связь? Какие показатели учитывают доход населения, занятость, образование, здравоохранение, преступность и другие характеристики качества жизни?

Далее при помощи кластерного анализа учащиеся выделяют несколько групп схожих территорий и визуализируют полученные результаты. В конечном счете учащиеся исследуют феномен культуры посредством формирования собственного индекса. Так или иначе, для того чтобы сформировать подобный индекс и ответить на поставленные вопросы, необходимо ознакомиться с самим феноменом

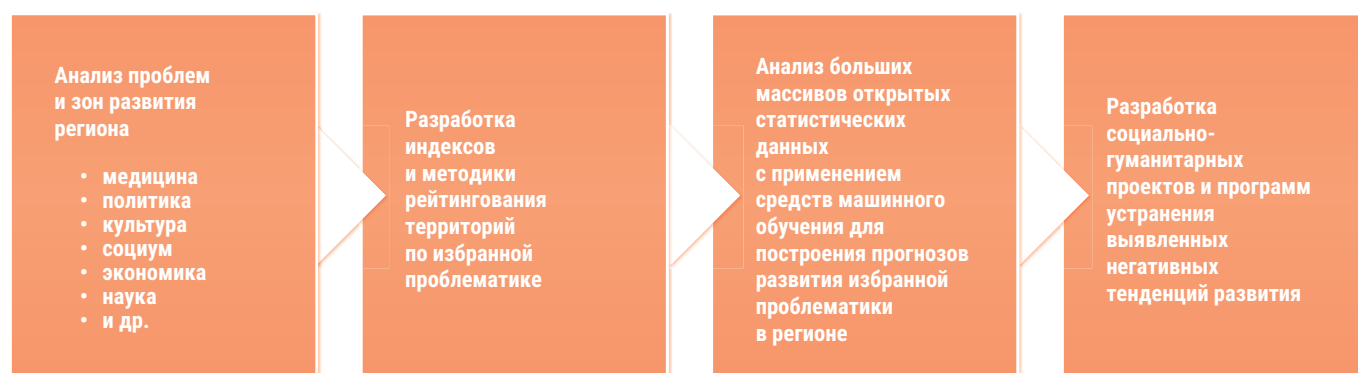
и использовать тот содержательный материал по данной тематике, который был получен в рамках основного образовательного процесса.

Допобразование как пристанище дата-грамотности

Сегодня не так часто можно встретить программы по формированию и развитию дата-грамотности в обозначенном нами ранее подходе (где наука о данных понимается в широком контексте применения). Те примеры, которые затрагивают общий контур обучения науке о данных детей подросткового и юношеского возраста, имеют разную продолжительность, а также в них всех присутствует разброс содержания (Srikant & Aggarwa, 2017; Bryant et al., 2019; Dryer et al., 2018; Datta et al., 2018; Mariescu-Istodor et al., 2019; Wolff et al., 2019). Наиболее приемлемым сегодня представляется реализация подобных программ в формате дополнительного образования, которое в большей степени ориентируется на интересы учащихся и более свободно с точки зрения требований при разработке самих учебно-тематических планов.

В качестве примера может выступать реализуемая коллективом авторов настоящей статьи программа «Дата Кампус», которая удерживает базовый фокус на погружении в гуманитарную проблематику и исследование региона проживания учащихся посредством применения науки о данных с попутным обучением языкам программирования. «Дата Кампус» представляет собой 10-дневную интенсивную сессию-погружение, что также является наиболее рациональным способом реализации подобных программ на данный момент. На сессии учащийся совершает полный цикл по поиску, обработке и интерпретации реальных данных, которые представляют собой региональную статистику по тематическому направлению в рамках интереса учащегося.

Попадая на «Дата Кампус», участники далеко не сразу принимают программу и анализировать большие массивы данных, как бы это парадоксально ни звучало. Прежде всего перед учащимся ставится вопрос о том, какие проблемы регионального развития его беспокоят. В ряду таких проблем могут быть совершенно разные: социальные, демографические,



медицинские, экологические, политические и так далее. Первое время, через форматы дискуссий и аналитических сессий, участники разрабатывают проблематику интересующих их сфер жизни и деятельности, работают с разнообразными аналитическими региональными материалами по рассматриваемой проблематике, а также разрабатывают методологию оценки (индекс), при помощи которой они смогут оценить в разных территориях уровень остроты проблематики. А далее, для того чтобы доказать степень выраженности проблемы, у ребят возникает необходимость работы со статистическими данными. И только на этом этапе начинается изучение инструментов программирования и работы с большими данными, так как они выступают ключом и средством анализа в рамках более объемных и волнующих ребят проблем. Становится понятно, зачем изучать науку о данных, как она поможет в решении конкретных вопросов.

Почему именно такая логика?

Обычно учащиеся (не только дети) осваивают науку о данных на абстрактных примерах в виде подготовленных массивов данных. Их значение и содержание не важны. Это делает обучение внеконтекстуальным и лишь создает иллюзию прикладной ценности осваиваемых навыков.

Проблемность такого подхода легко понять на примере изучения иностранных языков. Ведь мы осваиваем иностранный язык для того, чтобы его просто знать. Мы делаем это для того, чтобы разговаривать с другими людьми, читать необходимые тексты. Язык достаточно просто выучить, когда мы понимаем конкретные задачи, для решения которых он необходим. «Дата Кампус» придерживается подобной логики в отношении науки о больших

данных и программировании. В нашем случае изучать большие данные эффективнее, если они помогают понять конкретную проблему региона, в котором проживает ученик. И наоборот — ученику проще понять проблемы региона, если он использует средства анализа больших данных, так как в результате он получает не абстрактный ответ-предположение, а обнаруживает проблему в ее числовых значениях, получает возможность построить прогнозы и выявить закономерности на материале данных чисел. Такой подход *формирует и развивает дата-грамотность, которая позволяет применять инструменты анализа данных в совершенно разных областях жизни и деятельности.*

Литература

1. Anderson, P. et al. An undergraduate degree in data science: Curriculum and a decade of implementation experience // SIGCSE 2014 — Proceedings of the 45th ACM Technical Symposium on Computer Science Education. 2014. P. 145–150.
2. Baumer, B. S. A data science course for undergraduates: Thinking with data // The American Statistician. 2015. No. 69. P. 334–342.
3. Bryant, C. et al. A middle-school camp emphasizing data science and computing for social good // SIGCSE 2019 — Proceedings of the 50th ACM Technical Symposium on Computer Science Education. 2019. P. 358–364. DOI: 10.1145/3287324.3287510.
4. Bonaventura, M., Ciotti, V. et al. Predicting success in the worldwide startup network. 2019.
5. Cirillo, P., Taleb, N. The Decline of Violent Conflicts: What Do The Data Really Say? SSRN Electronic Journal. 2016. DOI: 10.2139/ssrn.2876315.
6. Cirillo, P., Taleb, N. What are the chances of a third world war? Real World Risk Institute Working Paper Series. 2016. DOI: 10.1111/j.1740-9713.2016.00903.x.
7. Datta, S., Nagabandi, V. Integrating data science and R programming at an early stage // IEEE 4th International Conference on Soft

Computing and Machine Intelligence, ISCM 2017. 2018. P. 1–5.

DOI: 10.1109/ISCM.2017.8279587.

8. De Veaux, R., Curriculum guidelines for undergraduate programs in data science // *Annual Review of Statistics and Its Application*. 2017. Vol. 4. No 1. P. 15–30.
9. Dryer, A., Walia, N., Chattopadhyay, A. A middle-school module for introducing data-mining, big data, ethics and privacy using rapidminer and a Hollywood theme // *SIGCSE 2018 – Proceedings of the 49th ACM Technical Symposium on Computer Science Education*. 2018. P. 753–758. DOI: 10.1145/3159450.3159553.
10. Finzer, W. The data science education dilemma // *Technology Innovations in Statistics Education*. 2013. Vol. 7. No. 2.
11. Gibson, J. P., Mourad T. The growing importance of data literacy in life science education // *American Journal of Botany*. 2018. Vol. 105. No 12. P. 1–4.
12. Heinemann, B. et al. Drafting a data science curriculum for secondary schools // *ACM International Conference Proceeding Series*. 2018. P. 1–5. DOI: 10.1145/3279720.3279737
13. Janosov, M., Muscioto, F. et al. Elites, communities and the limited benefits of mentorship in electronic music. 2019. DOI: <https://doi.org/10.1038/s41598-020-60055-w>.
14. Mariescu-Istodor, R., Jormanainen, I. Machine Learning Exercises for High School Students // *Proceedings of the 19th Koli Calling International Conference on Computing Education Research*. 2019.
15. Navarro, H., Miritello, G., Canales, A. et al. Temporal patterns behind the strength of persistent ties. *EPJ Data Sci*. 2017. Vol. 6. No 31. DOI: <https://doi.org/10.1140/epjds/s13688-017-0127-3>.
16. Schield, M. Information literacy, statistical literacy and data literacy // *IASSIST Quarterly*. 2018. Vol. 28. No 2. P. 6–11.
17. Srikant, S., & Aggarwal, V. (2017). Introducing data science to school kids // *Proceedings of the Conference on Integrating Technology into Computer Science Education, ITICSE*. P. 561–566. DOI: 10.1145/3017680.3017717.
18. Pangrazio, L., Sefton-Green, J. The social utility of data literacy // *Learning, Media and Technology*, 2019. DOI: 10.1080/17439884.2020.1707223.
19. Wolff, A., Wermelinger, M., Petre, M. Exploring design principles for data literacy activities to support children's inquiries from complex data // *International Journal of Human Computer Studies*. 2019. No 129 (March). P. 41–54. DOI: 10.1016/j.ijhcs.2019.03.006.

DATA LITERACY

A popular topic or a necessity?

Pavel P. GLUKHOV.

Research associate of the research sector “Open Education” of The Research Center for Socialization and Personalization of Children's Education (Federal Education Development Institute of The Russian Presidential Academy of National Economy and Public Administration). Expert at the laboratory of competence practices of education of the Moscow State Pedagogical University (9/1, Chernyakhovskogo ul., Moscow, 125319, Russian Federation). E-mail: gluhovpav.pav@gmail.com

Andrey A. DERYABIN.

MSc Social Psychology. Research associate of The Research Center for Socialization and Personalization of Children's Education (Federal Education Development Institute of The Russian Presidential Academy of National Economy and Public Administration) (9/1, Chernyakhovskogo ul., Moscow, 125319, Russian Federation). E-mail: deryabin-aa@ranepa.ru

Alexandre A. POPOV.

Doctor of Philosophy. Assistant professor, Head of The Research Center for Socialization and Personalization of Children's Education (Federal Education Development Institute of The Russian Presidential Academy of National Economy and Public Administration). Head of the laboratory of competence practices of the Institute of system projects of the Moscow State Pedagogical University. Professor of sociology and mass communications of the faculty of Humanities of the Novosibirsk State Technical University (9/1, Chernyakhovskogo ul., Moscow, 125319, Russian Federation). E-mail: aktor@mail.ru

Abstract

The article deals with the concept of data literacy. The article examines the content and composition of this type of literacy, justifies its relevance as an educational result consistent with the trends of digitalization. Approaches to the in-depth and General study of data science are divided. The article describes various types of tasks set on different educational materials and aimed at developing data literacy among students. We consider ways to deploy programs aimed at students “mastering data science without the need to form it into a separate discipline.

Keywords: data science, machine learning, educational programs, data literature, data analysis, Data Science, General education, career guidance, additional education, computer science.

References

- Anderson, P. et al. (2014). An undergraduate degree in data science: Curriculum and a decade of implementation experience. *SIGCSE 2014* (pp.145–150).
- Baumer, B. S. (2015). A data science course for undergraduates: Thinking with data. *The American Statistician*, 69, 334–342.
- Bonaventura, M., & Ciotti, V. et al. (2019). Predicting success in the worldwide start-up network.
- Bryant, C. et al. (2019). A middle-school camp emphasizing data science and computing for social good. *SIGCSE 2019 – Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 358–364). DOI:10.1145/3287324.3287510.
- Cirillo, P., & Taleb, N. (2016a). The Decline of Violent Conflicts: What Do The Data Really Say? *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2876315.

- Cirillo, P., & Taleb, N. (2016). What are the chances of a third world war? *Real World Risk Institute Working Paper Series*. DOI: 10.1111/j.1740-9713.2016.00903.x.
- Datta, S., & Nagabandi, V. (2018). Integrating data science and R programming at an early stage. *IEEE 4th International Conference on Soft Computing and Machine Intelligence, ISCMi 2017* (pp. 1–5). DOI:10.1109/ISCMi.2017.8279587.
- De Veaux, R. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application*, 4(1), 15–30.
- Dryer A., Walia N., & Chattopadhyay A. (2018). A middle-school module for introducing data-mining, big-data, ethics and privacy using rapidminer and a Hollywood theme. *SIGCSE 2018 – Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (pp. 753–758). DOI:10.1145/3159450.3159553.
- Finzer, W. (2013). The data science education dilemma. *Technology Innovations in Statistics Education*, 7(2).
- Gibson, J. P., & Mourad, T. (2018). The growing importance of data literacy in life science education. *American Journal of Botany*, 105(12), 1–4.
- Heinemann, B. et al. (2018). Drafting a data science curriculum for secondary schools. *ACM International Conference Proceeding Series*, 1–5. DOI:10.1145/3279720.3279737.
- Janosov, M., & Muscioto, F. et al. (2019). Elites, communities and the limited benefits of mentorship in electronic music. DOI: <https://doi.org/10.1038/s41598-020-60055-w>.
- Mariescu-Istodor, R., & Jormanainen, I. (2019). *Machine Learning Exercises for High School Students. Proceedings of the 19th Koli Calling International Conference on Computing Education Research*. Retrieved from http://www.cs.columbia.edu/~CS4HS/talks/ml_for_hs.pdf.
- Navarro, H., Miritello, G., & Canales, A. et al. (2017). Temporal patterns behind the strength of persistent ties. *EPJ Data Sci.*, 6(31). DOI: <https://doi.org/10.1140/epjds/s13688-017-0127-3>.
- Pangrazio, L., & Sefton-Green, J. (2019). The social utility of data literac. *Learning, Media and Technology*. DOI: 10.1080/17439884.2020.1707223.
- Schild, M. (2018). Information literacy, statistical literacy and data literacy. *IASSIST Quarterly*, 28(2), 6–11.
- Srikant, S., & Aggarwal, V. (2017). Introducing data science to school kids. *Proceedings of the Conference on Integrating Technology into Computer Science Education, ITiCSE* (pp. 561–566). DOI: 10.1145/3017680.3017717.
- Wolff, A., Wermelinger, M., & Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human Computer Studies*, 129 (March), 41–54. DOI: 10.1016/j.ijhcs.2019.03.006.