

ИНФОРМАЦИОННОЕ ОБЩЕСТВО

1

2021

Цифра открывает любую дверь

Как блогеры влияют на покупателей

Инновационные инструменты для взаимодействия органов власти с гражданами

Образовательные программы в области анализа данных и машинного обучения

Единое цифровое пространство научных знаний

Информационные права человека

Новый вид бизнеса: купля-продажа персональных данных

Экономические модели на основе анализа больших данных

№ 1
2021

ИНФОРМАЦИОННОЕ ОБЩЕСТВО

ОСНОВАН В 1989 ГОДУ
ВЫХОДИТ 6 РАЗ В ГОД

УЧРЕДИТЕЛИ:

ИНСТИТУТ РАЗВИТИЯ ИНФОРМАЦИОННОГО ОБЩЕСТВА
РОССИЙСКАЯ ИНЖЕНЕРНАЯ АКАДЕМИЯ

ГЛАВНЫЙ РЕДАКТОР: РЕДАКЦИОННЫЙ СОВЕТ:

ЕРШОВА Татьяна
Викторовна – канд.
экон. наук

ХОХЛОВ Юрий Евгеньевич (председатель) – канд. физ.–мат. наук, доц., acad. РИА
ОРЛОВ Степан Владимирович (зам. председателя) – канд. экон. наук
АЛЕКСЕЕВА Ирина Юрьевна – д-р филос. наук, доц.
БОГДАНОВ Александр Владимирович – д-р физ.–мат. наук, проф.
ВАРТАНОВА Елена Леонидовна – д-р фил. наук, проф.
ВЕРШИНСКАЯ Ольга Николаевна – д-р экон. наук
ВОЙСКУНСКИЙ Александр Евгеньевич – д-р психол. наук
ДЕЖИНА Ирина Геннадьевна – д-р экон. наук, проф.
ЕЛИЗАРОВ Александр Михайлович – д-р физ.–мат. наук, проф.
ЗАСУРСКИЙ Ясен Николаевич – д-р фил. наук, проф.
ИВАНОВ Алексей Дмитриевич – д-р экон. наук, чл.-кор. РАЕН
ИВАХНЕНКО Евгений Николаевич – д-р филос. наук, проф.
КОГАЛОВСКИЙ Михаил Рувимович – канд. техн. наук, доц.
КОЛИН Константин Константинович – д-р техн. наук, проф., засл. деятель науки РФ
КУЗНЕЦОВА Наталия Ивановна – д-р филос. наук, проф.
МЕНДЖКОВИЧ Андрей Семенович – д-р химических наук, ст. науч. сотрудник
ОЛЕЙНИК Андрей Владимирович – д-р техн. наук, проф.
РАЙКОВ Александр Николаевич – д-р техн. наук, проф.
РУСАКОВ Александр Ильич – д-р хим. наук, проф.
СЕМЕНОВ Алексей Львович – д-р физ.–мат. наук, acad. РАН, действ. член РАО
СЕМЕНОВ Евгений Васильевич – д-р филос. наук, проф.
СЕРДЮК Владимир Александрович – канд. техн. наук, доц.
СМОЛЯН Георгий Львович – д-р филос. наук, проф.
СТРЕЛЬЦОВ Анатолий Александрович – д-р техн. наук, д-р юрид. наук, проф., засл. деятель науки РФ
ТАТАРОВА Галина Галеевна – д-р социол. наук, проф.
ЧЕРЕШКИН Дмитрий Семенович – д-р техн. наук, проф., acad. РАЕН
ШАПОШНИК Сергей Борисович
ЩУР Лев Николаевич – д-р физ.–мат. наук, проф.
ЯКУШЕВ Михаил Владимирович

Журнал зарегистрирован в Роспечати
(Per № 015 766 от 01.07.1999)
ISSN 1605-9921 (эл.)

Адрес редакции: Москва, Армянский переулок,
д. 9, офис 310
Для переписки: 101000, Москва, Главпочтамт, а/я 716
Тел.: +7 (495) 912-22-29
Электронная почта: info@infosoc.iis.ru
Веб-сайт: www.infosoc.iis.ru

Позиция редакции может не совпадать с мнением авторов.

Авторы несут ответственность за патентную чистоту, достоверность и точность приведенных фактов, цитат, экономико-статистических данных, собственных имен, географических названий и прочих сведений, а также за разглашение данных, не подлежащих открытой публикации. При любом использовании оригинальных материалов ссылка на журнал обязательна.

ЛЕГАЛЬНЫЙ ПОЛЬЗОВАТЕЛЬ
Пара(-)Тайп
IN LEGAL USE

В макете журнала использованы шрифты
ООО НПЦ «ПараТайп»

ПУБЛИКУЕМЫЕ МАТЕРИАЛЫ ПРОШЛИ ПРОЦЕДУРУ
РЕЦЕНЗИРОВАНИЯ И ЭКСПЕРТНОГО ОТБОРА

ЖУРНАЛ ВКЛЮЧЕН В ПЕРЕЧЕНЬ ПЕРИОДИЧЕСКИХ ИЗДАНИЙ, РЕКОМЕНДОВАННЫХ ВЫСШЕЙ АТТЕСТАЦИОННОЙ КОМИССИЕЙ
РФ ДЛЯ ПУБЛИКАЦИИ МАТЕРИАЛОВ КАНДИДАТСКИХ И ДОКТОРСКИХ ДИССЕРТАЦИОННЫХ ИССЛЕДОВАНИЙ. ЖУРНАЛ ВХОДИТ В
ДАННЫЙ СПИСОК С 26 ФЕВРАЛЯ 2010 ГОДА.
С 2015 ГОДА ЖУРНАЛ ВХОДИТ В РОССИЙСКУЮ ПОЛКУ ЖУРНАЛОВ (RUSSIAN SCIENCE CITATION INDEX) НА ПЛАТФОРМЕ WEB OF
SCIENCE.

© Институт развития информационного общества, 2021

Публикации в журнале «Информационное общество» доступны в открытом доступе по международной лицензии
Creative Commons «С указанием авторства - Некоммерческая - С сохранением условий» версии 4.0 Международная

СОДЕРЖАНИЕ № 1 2021

СЛОВО ГЛАВНОГО РЕДАКТОРА

- 1 ЕРШОВА Татьяна Викторовна **Alea jacta est**

ЦИФРОВАЯ ЭКОНОМИКА

- 2 АЛЕКСАНДРОВСКИЙ Сергей Владимирович, СИЛКИНА Ольга Сергеевна **Как лидеры мнений помогают брендам на рынке общественного питания**

ИНФОРМАЦИОННОЕ ОБЩЕСТВО И ВЛАСТЬ

- 15 ПОГОДИНА Ирина Владимировна, АВДЕЕВ Данила Алексеевич **Инструменты электронного участия граждан в публичном управлении: понятие и классификация**

ОБРАЗОВАНИЕ В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ

- 21 ДЕРЯБИН Андрей Александрович, ПОПОВ Александр Анатольевич **Обучение дата-грамотности: содержание и конкурентные преимущества образовательной программы**

НАУКА И ИННОВАЦИИ В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ

- 30 КАЛЕНОВ Николай Евгеньевич, СОБОЛЕВСКАЯ Ирина Николаевна, СОТНИКОВ Александр Николаевич **Единое цифровое пространство научных знаний в мировом информационном пространстве**

ИНФОРМАЦИОННОЕ ОБЩЕСТВО И ПРАВО

- 42 ЛЯСКОВСКИЙ Иван Иванович **Содержание права на информацию в контексте режима информации открытого и ограниченного доступа**

ДОВЕРИЕ И БЕЗОПАСНОСТЬ В ИНФОРМАЦИОННОМ ОБЩЕСТВЕ

- 49 АЛИГУЛИЕВ Расим Магамед оглу, МАХМУДОВ Расим Шариф оглу **Особенности «черного рынка» персональных данных и создаваемые им проблемы**

ТЕХНОЛОГИИ ИНФОРМАЦИОННОГО ОБЩЕСТВА

- 56 ДОРЖИЕВ Ардан Саянович **Прогнозирование банкротства методами машинного обучения**

Слово главного редактора

ALEA JACTA EST

Ершова Татьяна Викторовна

Кандидат экономических наук

Институт развития информационного общества, генеральный директор

Научно-аналитический журнал «Информационное общество», главный редактор

Член Союза журналистов России

Член Международной федерации журналистов

Москва, Российская Федерация

info@infosoc.iis.ru

В теме первого номера 2021 года, а именно «Цифра открывает любую дверь», многие узнают перефразированное латинское крылатое выражение “Auro quaque janua panditur” (золото открывает любую дверь). Поистине, уже не деньги, а цифровые технологии становятся всемогущими и создают безграничные возможности. Ну разве это не чудо, когда ты за пару часов сам можешь спланировать потрясающую десятидневную туристическую поездку, охватив главные достопримечательности довольно большого региона, выбрав подходящие рейсы и удобные места в самолете, забронировав две гостиницы и три экскурсии и даже записавшись на экзотическую зимнюю спа-процедуру прямо на берегу моря?! И сразу оплатив все означенное!

Это всего лишь маленький пример из жизни обычного современного человека, вооруженного любознательностью, здравым смыслом (дабы не нарваться на жуликов) и бытовыми вычислительными устройствами. Что уж там говорить о возможностях квантовых компьютеров, цифровых двойников и искусственного интеллекта для производства, науки, культуры, торговли, управления? А сколько всего принесут с собой всякие новые виды вычислений – облачных, туманных, росистых... От красоты метафор и невиданных перспектив просто дух захватывает!

Прогрессивное человечество некоторое время назад, как и Гай Юлий Цезарь в 49 году до нашей эры, сказало: “Alea jacta est” (жребий брошен) и пересекло свой Рубикон – границу между индустриальным и информационным обществом. В словах Цезаря о жребии был заключен глубокий смысл: он осознавал необратимость происходящего и действительно не знал, что ждет его впереди. Но если великий римлянин, идя на риск ради высокой цели, уповал только на судьбу, мы можем полагаться уже на более надежные вещи: опыт лидеров и научные достижения. Мы в нашем журнале стараемся отражать и то, и другое, и это очень важная часть нашей миссии. Другая ее часть – популяризировать новые технологии и их приложения, просвещать аудиторию по вопросам квалифицированного и безопасного их использования, помогать преодолевать естественный страх перед неизведанным.

В конце прошлого года, анализируя нашу публикационную активность, мы самокритично отметили, что некоторые важные тематические рубрики оставались без должного внимания. В первом номере журнала за этот год мы постарались частично преодолеть этот недостаток, разместив статьи в рубриках «Наука и инновации в информационном обществе» и «Информационное общество и право». Другие статьи распределились по традиционно популярным рубрикам: «Цифровая экономика», «Информационное общество и власть», «Образование в информационном обществе», «Доверие и безопасность в информационном обществе», «Технологии информационного общества».

Нас очень радует интерес к нашему журналу ученых и исследователей из других стран. В этом номере мы публикуем статьи авторов из Азербайджана и Беларуси. Материалы из России представлены авторами из Москвы и Владимира. Мы очень надеемся, что географическое разнообразие нашего редакционного портфеля будет со временем только расти.

Как вы заметили, уважаемые читатели, наш журнал стал выходить строго по графику, без единого дня задержки. Это стало возможным благодаря переходу на цифровую платформу: полагаясь только на себя и своих авторов, мы перестали зависеть от доступности внешних специалистов, трудностей типографии, превратностей логистики. Процесс публикации стал проще и дешевле, ценные научные материалы – доступнее, взаимодействие с авторами – менее громоздким. Как тут не воспеть оду цифровым технологиям и не призвать коллег-издателей воспользоваться нашим опытом!

© Ершова Т.В., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

Цифровая экономика**КАК ЛИДЕРЫ МНЕНИЙ ПОМОГАЮТ БРЕНДАМ НА РЫНКЕ
ОБЩЕСТВЕННОГО ПИТАНИЯ**

Статья рекомендована к публикации членом редакционного совета И.Ю. Алексеевой 20.12.2020.

Александровский Сергей Владимирович

К.э.н.

*Национальный исследовательский университет «Высшая школа экономики», доцент
Россия, Москва*

saleksandrovskiy@hse.ru

Силкина Ольга Сергеевна

*Национальный исследовательский университет «Высшая школа экономики», доцент
Россия, Москва*

silkinaolgann@yandex.ru

Аннотация

Инстаграм – одна из самых популярных онлайн-платформ у российских пользователей. При выборе товаров и услуг потребители спрашивают рекомендации у друзей, поэтому дружеские рекомендации от лидеров мнений в Инстаграм оказывают влияние на решение о покупке. Компании оценивают эффективность рекламы у лидеров мнений по вовлеченности пользователей в рекламные посты, охвату и стоимости контакта. На основе эмпирического исследования брендов на рынке общественного питания Санкт-Петербурга авторы выяснили, что уровень вовлеченности пользователей в рекламный пост у крупных блогеров ниже, чем у мелких блогеров, средняя стоимость охвата ниже у крупных лидеров мнений, лайфстайл-блогеров и у блогеров с уровнем вовлеченности больше 2%.

Ключевые слова

Инстаграм, лидер мнений, блогер, продвижение, реклама, вовлеченность, стоимость охвата

Введение

Потребители устали от рекламы. Программы, блокирующие рекламу, стоят на компьютерах у 27% пользователей [6], но в то же время потребители положительно относятся к рекомендациям лидеров мнений (opinion leader или influencer) в интернете, читают блоги, комментарии под постами в социальных сетях.

Потребители больше доверяют рекомендациям лидеров мнений, чем традиционной рекламе [4, 7]. Доверие растет, если предыдущие рекомендации лидера мнений оказались полезными [14].

Реклама у лидеров мнений положительно влияет на повышение узнаваемости бренда [4], отношение потребителей к рекламируемому бренду и доверие бренду [11, 25], намерение совершить покупку [14] и поведение потребителей [13]. Для российских компаний продвижение через лидеров мнений также становится важным рекламным каналом.

В 2018 году в России на онлайн-рекламу компании потратили 203 млрд руб., в т.ч. на рекламу у блогеров – 5 млрд руб. В 2019 году эксперты прогнозировали рост расходов на рекламу у блогеров до 10 млрд руб. [29].

В мире в первой половине 2020 года расходы на онлайн рекламу снизились на 30% в условиях пандемии. Это меньше, чем сокращение размещений в традиционных каналах. Рекламные бюджеты остаются в онлайн [8]. Во время пандемии 2020 года рекламодатели увеличили свои расходы на рекламу у блогеров на 22% [22].

© Александровский С.В., Силкина О.С., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

Инстаграм входит в топ-3 онлайн-платформ по популярности у российских пользователей и уступает только ВКонтакте и YouTube [15]. При выборе товаров и услуг потребители спрашивают рекомендации у друзей, поэтому дружеские рекомендации от лидеров мнений (блогеров) в Инстаграм оказывают большое влияние на решение потребителей о покупке. Россия на первом месте среди европейских стран по числу активных пользователей Инстаграм. Продвижение в Инстаграм через лидеров мнений – это один из самых быстрорастущих каналов рекламы в России. [3, 21].

Компании оценивают эффективность размещения рекламы у лидеров мнений по вовлеченности пользователей в рекламные посты, охвату постов и стоимости контакта. На решение компании о выборе лидера мнений для размещения рекламы влияет количество подписчиков у лидера мнений, стоимость размещения поста и специализация блогера.

Количество подписчиков – главный критерий выбора для размещения рекламы. Количество подписчиков – это индикатор влияния блогера [17]. Потребители больше доверяют лидерам мнений с большим числом подписчиков, т.е. больше доверяют и рекламе, которую размещают у себя лидеры мнений с большим числом подписчиков [16]. Компании стремятся разместить рекламу у блогеров с большей аудиторией, чтобы получить больший охват для рекламы, но действительно размер аудитории определяет эффективность рекламы?

Цель данного исследования – проверить влияние числа подписчиков и тематики блогера на эффективность рекламы у лидеров мнений в Инстаграм. Авторы отвечают на вопрос: как влияет количество подписчиков в аккаунте лидера мнений и тематика блогера на эффективность рекламы бренда (вовлеченность в рекламный пост, охват и стоимость охвата) у лидера мнений?

Научная новизна заключается в том, что авторы на основе эмпирического исследования проверяют устоявшиеся мнения об эффективности рекламы у блогеров и показывают новые зависимости критериев выбора блогера и эффективности рекламы.

1 Понятия лидер мнений и блогер

В зарубежных работах по исследованию социальных сетей (Инстаграм, Фейсбук, Ютуб и др.) авторы используют понятия «лидер мнений» и «блогер». Лидеры мнений существуют на всех ступенях общества и во всех социально-демографических группах [30]. Лидеров мнения отличают активная жизненная позиция, разветвлённая сеть социальных контактов, стремление распространять полезную информацию, уверенность при формировании тренда и вовлеченность в процесс решения проблем [2, 12]. Лидеров мнений онлайн отличает вовлеченность. Блог – это сетевой дневник одного или нескольких авторов, состоящий из записей в обратном хронологическом порядке [18, 27]. Блогер – это человек или группа лиц, которые ведут блог в Интернете. Лидером мнений называют популярного человека как онлайн, так и офлайн. Блогер получил популярность именно в интернете благодаря ведению блога, т.е. блогер – это лидер мнений на определенной площадке в интернете [31]. В зарубежных работах авторы используют понятия «лидер мнений» и «блогер» как синонимы. В данной работе авторы используют эти понятия как равнозначные.

Исходя из характеристик блогера, компании подбирают подходящий формат взаимодействия: рекламный пост, stories, амбассадорство, прямой эфир, интеграции в видео и др. [10]. В данном исследовании авторы оценивают эффективность рекламного поста как наиболее распространенного формата рекламы у лидера мнений. В отличие от stories, пост сохраняется в ленте Инстаграм в течение долгого времени, поэтому авторы смогли собрать больше данных, а пользователи с большей вероятностью увидели рекламу.

Рекламодатели оценивают эффективность рекламы у лидеров мнений так же, как и в других рекламных каналах на основе статистики и метрик. Лидеры мнений предоставляют статистику рекламных кампаний с площадки (канала). Агентства и рекламодатели оценивают результаты с помощью реализованных промокодов, регистраций чеков, динамики подписчиков аккаунта бренда до и после рекламы, количества участников в конкурсе и др. метрик [9].

2 Методология эмпирического исследования

Для исследования эффективности рекламы у Инстаграм-блогеров авторы выбрали десять заведений общественного питания Санкт-Петербурга (далее – брендов): «Макаронники», «Pita's»,

«Yami Yami», «Ешь фреш», «Пышечка», «Pinch!», «Italy», «Краснодарский парень», «Starbucks» и «Tsunami» (см. табл.1).

Авторы составили описание брендов на основе отзывов [19, 23], в которых клиенты чаще всего оценивали бренд по среднему чеку, кухне и концепции.

Все бренды размещают рекламу у блогеров в Инстаграм. Авторы включили в выборку бренды в разных ценовых сегментах и с разными концепциями. Такая выборка брендов позволит авторам обобщить выводы исследования о продвижении через лидеров мнений¹ в Инстаграм для широкого круга компаний (брендов) на рынке общественного питания.

Таблица 1. Бренды, участвующие в исследовании

Бренд	Концепция	Кухня	Ср. чек, руб.
Макаронники	Ресторан	Итальянская	1700
Pita's	Стритфуд-кафе	Блюда в тортильях с начинками: шаверма, бурито, фалафель	300
Yami Yami	Доставка еды	Азиатская кухня	700
Ешь фреш	Кафе	Смузи и блюда сыроедческой кухни	300
Пышечка	Сеть пышечных	Пышки и блюда, ассоциирующиеся с советским временем	50
Pinch!	Тапас-бар	Испанская кухня: закуски (пинчос и тапас), супы, горячее	600
Italy	Сеть ресторанов	Итальянская, японская и паназиатская кухни	1400
Краснодарский парень	Бургерная	Бургеры, сэндвичи, салаты, легкие закуски	500
Starbucks	Сеть кофеен	Кофе, чай, десерты, выпечка, салаты и сэндвичи	550
Tsunami	Ресторан-бар	Японские закуски и азиатская кухня	1100

Авторы выбрали Санкт-Петербург, т.к. это крупный российский город, в котором продвижение брендов через блогеров развито лучше, чем в небольших городах. На примере рынка Санкт-Петербурга авторы проверили гипотезы об эффективности рекламы брендов у лидеров мнений.

На рынке общественного питания бренды чаще работают с блогерами до 100 тыс. подписчиков, чем с блогерами, у которых 100-500 тыс. подписчиков, или крупными лидерами мнений, у которых от 500 тыс. подписчиков и больше. Дружеская рекомендация бренда от мелкого блогера воспринимается аудиторией более доверительно, чем рекомендации от средних и крупных блогеров [16, 17].

Стоимость рекламного поста у крупных и средних блогеров выше, чем у мелких блогеров. Из-за ограниченного бюджета локальные российские бренды реже работают со средними и крупными блогерами. Международные бренды (Starbucks) в отличие от локальных выбирают дорогих крупных блогеров.

Бренды на рынке общественного питания часто размещают рекламу у блогеров в тематике лайфстайл², т.к. личная рекомендация бренда общественного питания органично вписывается в контент лайфстайл-блогера и вызывает больше доверия у аудитории.

Бренды, которые авторы выбрали для исследования, размещают рекламу у блогеров с числом подписчиков от 1 971 до 1,6 млн человек, т.е. охватывают блогеров всех размеров. В выборку включены 274 рекламных поста от лидеров мнений для 10 выбранных брендов.

Для оценки лидеров мнений авторы использовали ключевые показатели эффективности аккаунтов и рекламных постов от Interactive Advertising Bureau³ [9]:

¹ Через посты, которые лидеры мнений самостоятельно размещают в Инстаграм на своих страницах [31].

² Блогер, который создает контент на основе своих личных интересов и повседневной деятельности. Lifestyle-блоги очень персонализированы и рассказывают о местоположении, этапе жизни и опыте блогера [32,33].

- Количество подписчиков (followers);
- Уровень вовлеченности (engagement rate или er)⁴;
- Предполагаемая стоимость охвата 1000 человек (срм)⁵.

Рекламодатели размещают рекламу у лидеров мнений с небольшим количеством подписчиков, т.к. уверены, что аудитория этих блогеров не устала от рекламы, воспринимает блогера как друга и больше доверяет рекомендациям блогера. Зарубежные авторы утверждают, что сотрудничество с небольшими блогерами помогает брендам сохранять собственную уникальность [24]. Российские исследования подтверждают, что с ростом количества подписчиков у лидера мнений вовлеченность в рекламный пост лидера мнений падает, но чем больше подписчиков, тем дешевле охват аудитории [32, 33]. Это значит, что реклама у блогера может быть эффективна по одной метрике из списка IAB и неэффективна по другой.

У блогеров, с которыми работают бренды «Pita's», «Yami Yami», «Пышечка», «Tsunami», уровень вовлеченности выше среднего значения по выборке, а число подписчиков – ниже среднего значения по выборке (см. табл.2). Эти бренды ориентируются на показатель уровня вовлеченности для оценки качества блогера и выбирают мелких блогеров, что может влиять на средний уровень вовлеченности. Бренды «Макароны», «Starbucks», «Pinch!», которые выбирают более крупных блогеров, уступают им по вовлеченности.

Таблица 2. Сравнение брендов по эффективности рекламных постов, размещенных у лидеров мнений в Инстаграм

Бренд	Средний уровень вовлеченности	Средняя стоимость охвата 1000 человек, руб.	Среднее число подписчиков у лидера мнений
Starbucks	0,0409	179,97	266 396
Ешь фреш	0,0551	628,68	86 139
Pinch	0,0442	362,52	62 714
Макароны	0,0405	285,90	57 823
Pitas	0,1086	315,92	55 079
Tsunami	0,0576	251,60	52 872
Italy	0,0249	320,72	46 623
Пышечка	0,0590	429,08	42 722
Yami Yami	0,0725	557,19	37 614
Краснодарский парень	0,0509	379,35	32 086
Вся выборка	0,0562	364,76	73 737

Для дальнейшего анализа авторы разделили выборку на две группы по числу подписчиков у лидеров мнений, как это делают в зарубежных исследованиях [5] (см. табл.3):

- Мелкие блогеры, у которых до 100 тыс. подписчиков;
- Крупные блогеры, у которых свыше 100 тыс. подписчиков.

На основе своих наблюдений авторы выдвинули гипотезы 1а и 1б:

- (Гипотеза 1а) У крупных и мелких блогеров средний уровень вовлеченности в рекламный пост отличается.
- (Гипотеза 1б) У крупных блогеров уровень вовлеченности в рекламный пост ниже, чем у мелких блогеров.

³ Международная организация, которая разрабатывает отраслевые стандарты, проводит исследования и оказывает юридическую поддержку индустрии онлайн-рекламы.

⁴ Количество взаимодействий с контентом блогера (лайки, комментарии, сохранения), разделенное на количество подписчиков.

⁵ Показатель аналогичный CPM (cost per mile) для традиционных каналов. Рассчитан как предполагаемая стоимость одного контакта, умноженная на 1000. Предполагаемая стоимость одного контакта рассчитана как отношение предполагаемой стоимости размещения к предполагаемому охвату рекламного поста.

Таблица 3. Сравнение крупных и мелких блогеров по эффективности рекламных постов, размещенных в Инстаграм

Группы блогеров по количеству подписчиков	Средний уровень вовлеченности	Средняя стоимость охвата 1000 человек, руб.	Количество наблюдений
Мелкие блогеры (до 100 тыс. подписчиков)	0,05865426	392,42	223
Крупные блогеры (свыше 100 тыс. подписчиков)	0,045527451	243,79	51
Вся выборка	0,056210949	364,76	274

Крупные блогеры привлекают бренды быстрым и большим охватом, но стоимость размещения растет с ростом количества подписчиков [28]. Средняя стоимость охвата 1000 человек ниже среднего значения по выборке у таких брендов как «Starbucks», «Tsunami» и «Макароны», выше среднего - у «Yami Yami», «Ешь фреш», «Пышечка». «Ешь фреш» продвигает себя с помощью лидеров мнений с большим количеством подписчиков, чем у других брендов, «Tsunami» - у лидеров мнений с меньшим числом подписчиков.

На основе наблюдений авторы выдвинули гипотезы 2а и 2б:

- (Гипотеза 2а) У крупных и мелких блогеров средняя стоимость охвата 1000 человек в рекламном посте отличается.
- (Гипотеза 2б) У крупных блогеров средняя стоимость охвата 1000 человек в рекламном посте выше, чем у мелких блогеров.

Бренды на рынке общественного питания работают не только с лайфстайл-блогерами, но также с блогерами узконаправленных тематик: творчество, путешествия, красота, ЗОЖ и др. Авторы разделили выборку на группы по тематике: лайфстайл-блогеры и блогеры с узконаправленной тематикой (см. табл.4)

Таблица 4. Сравнение лайфстайл-блогеров с блогерами узконаправленных тематик по эффективности рекламных постов, размещенных в Инстаграм

Группы блогеров по тематике	Средний уровень вовлеченности	Средняя стоимость охвата 1000 человек, руб.	Количество наблюдений
Лайстайл	0,051168456	372,69	149
Узконаправленные	0,0622216	355,29	125
Вся выборка	0,056210949	364,75	274

На основе наблюдений авторы выдвинули гипотезы 3 и 4:

- (Гипотеза 3) У лайфстайл-блогеров уровень вовлеченности в рекламный пост выше, чем у блогеров с узконаправленной тематикой.
- (Гипотеза 4) У лайфстайл-блогеров стоимость охвата 1000 человек в рекламном посте ниже, чем у блогеров с узконаправленной тематикой.

Авторы также решили проверить значимость различий между средней стоимостью охвата 1000 человек в рекламном посте у блогеров с уровнем вовлеченности больше и меньше 2% (Табл. 5). Если у блогера уровень вовлеченности ниже 2%, то качество аудитории блогера вызывает сомнения, т.к. аудитория не вовлекается в контент блогера [32].

- (Гипотеза 5) У блогеров с уровнем вовлеченности в рекламный пост $\geq 2\%$ стоимость охвата 1000 человек ниже, чем у блогеров с уровнем вовлеченности $< 2\%$.

Табл. 5 Сравнение блогеров с разным уровнем вовлеченности по эффективности рекламных постов, размещенных в Инстаграм

Группы блогеров по вовлеченности	Средний уровень вовлеченности	Средняя стоимость охвата 1000 человек, руб.	Количество наблюдений
$\geq 2\%$	0,012325	602,23	56

Группы блогеров по вовлеченности	Средний уровень вовлеченности	Средняя стоимость охвата 1000 человек, руб.	Количество наблюдений
< 2 %	0,067484404	303,75	218
Вся выборка	0,056210949	364,75	274

3 Результаты

Перед проверкой гипотез авторы проверили данные на нормальность. Распределение данных по уровню вовлеченности, средней стоимости охвата 1000 человек отличается от нормального, см. рис. 1-3. Поэтому для сравнения средних между группами авторы использовали непараметрический тест групповых различий для независимых выборок – тест Вилкоксона, (или тест Манна-Уитни).

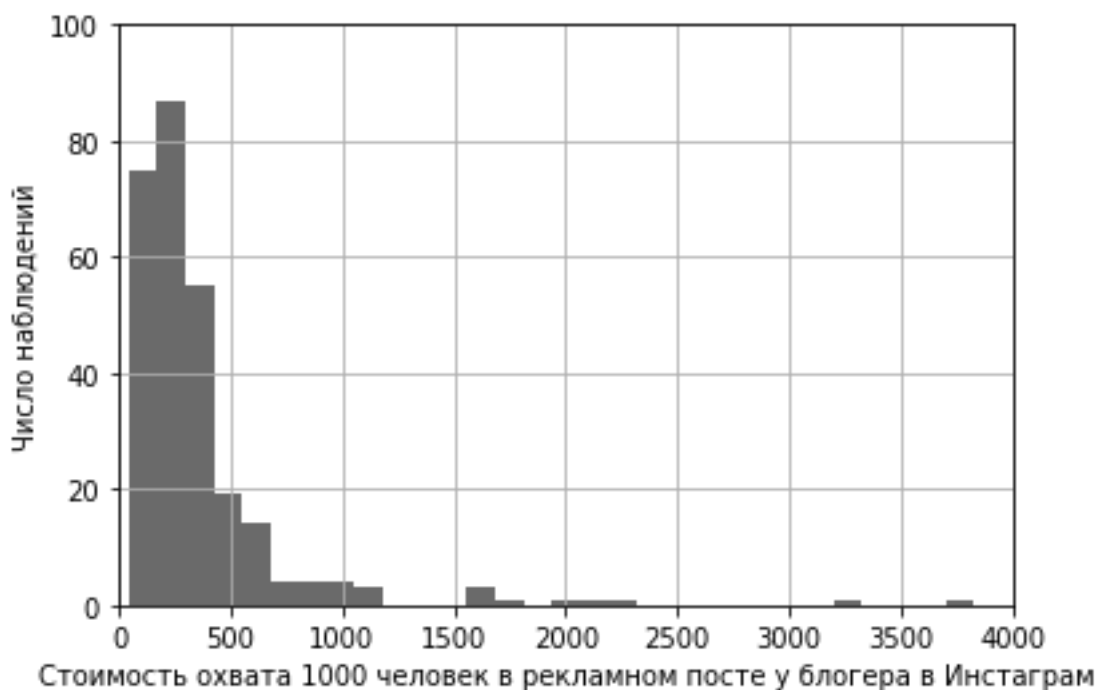


Рис. 1. Распределение стоимости охвата 1000 человек

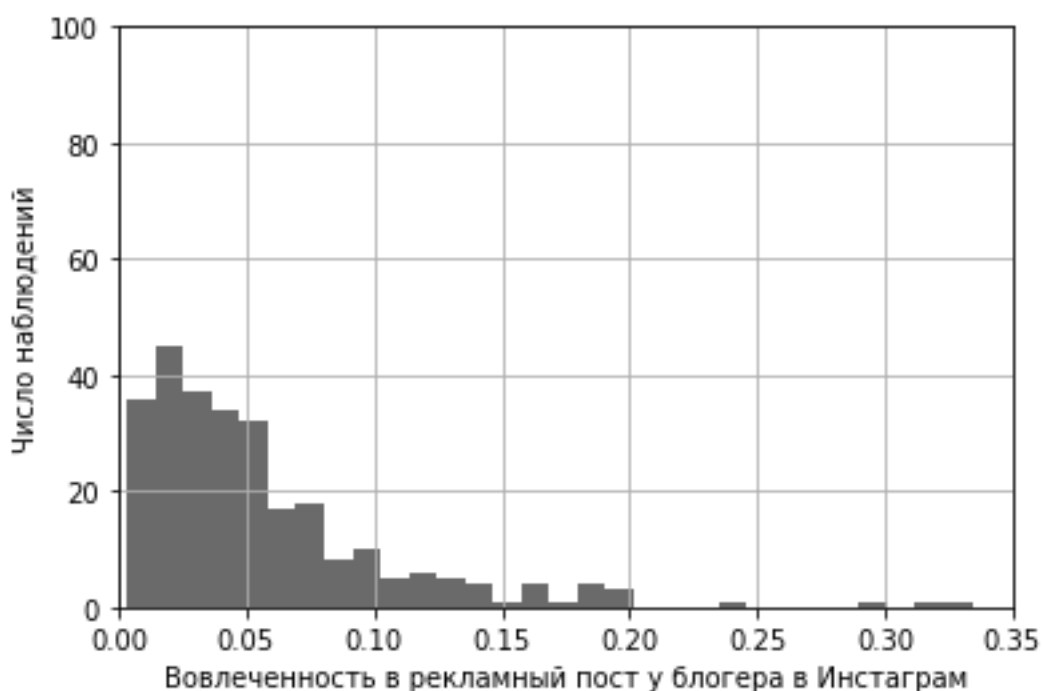


Рис. 2. Распределение уровня вовлеченности в рекламный пост у блогера в Инстаграм

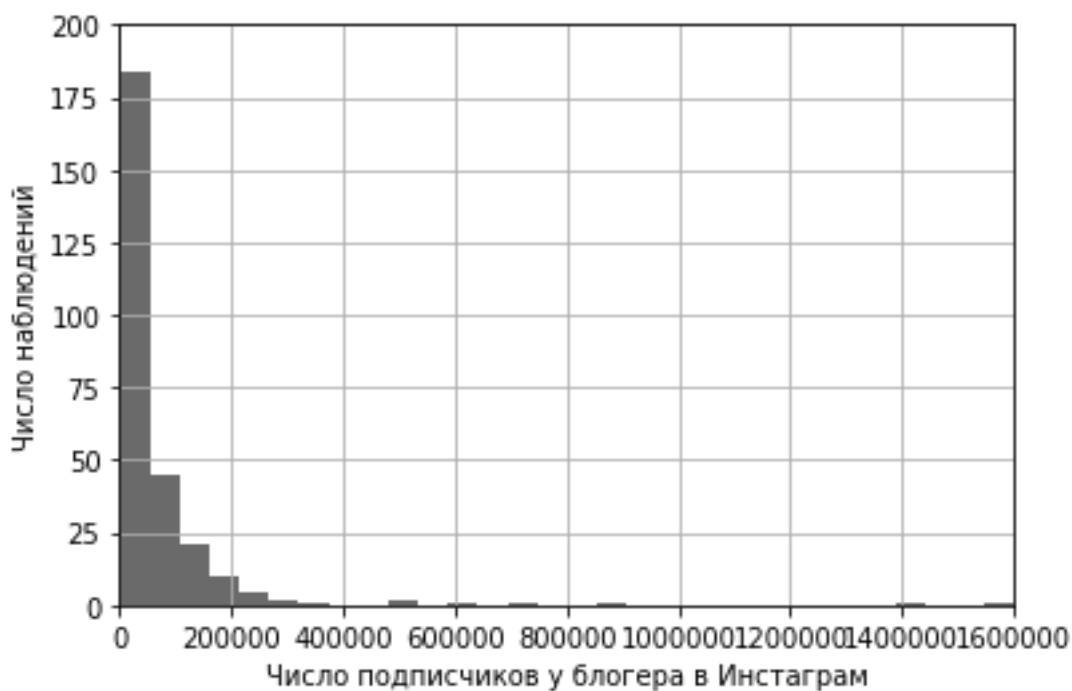


Рис. 3. Распределение числа подписчиков у блогера в Инстаграм

Гипотеза 1а подтверждена ($p\text{-value } 0.03165 < 0.05$). Существуют статистически значимые отличия в среднем уровне вовлеченности в рекламный пост у крупных и мелких блогеров.

Гипотеза 1б подтверждена, т.к. у крупных блогеров средняя вовлеченность в рекламный пост (0.04552745) ниже, чем у мелких блогеров (0.05865426).

Гипотеза 2а подтверждена ($p\text{-value } 1.956e-08 < 0.05$). Существуют статистически значимые отличия в средней стоимости охвата 1000 человек у крупных и мелких блогеров.

Гипотеза 2б отклонена, т.к. у крупных блогеров средняя стоимость охвата 1000 человек в рекламном посте (243.79 руб.) ниже, чем у мелких блогеров (392.42 руб.).

Гипотеза 3 отклонена ($p\text{-value } 0.06163 > 0.05$). Авторы не обнаружили статистически значимые отличия в среднем уровне вовлеченности в рекламный пост у лайфстайл-блогеров и блогеров с узконаправленной тематикой.

Гипотеза 4 подтверждена ($p\text{-value } 0.00435 < 0.05$). Существуют статистически значимые отличия в средней стоимости охвата 1000 человек в рекламном посте у лайфстайл-блогеров и блогеров с узконаправленной тематикой. У лайфстайл-блогеров средняя стоимость охвата 1000 человек (355.29 руб.) ниже, чем у блогеров с узконаправленной тематикой (372.7 руб.).

Гипотеза 5 подтверждена ($p\text{-value } 1.607e-05 < 0.05$). Существуют статистически значимые отличия в средней стоимости охвата 1000 человек в рекламном посте у блогеров с уровнем вовлеченности $\geq 2\%$ и $< 2\%$. У блогеров с уровнем вовлеченности $\geq 2\%$ средняя стоимость охвата 1000 человек (303.75 руб.) ниже, чем у блогеров с уровнем вовлеченности $< 2\%$ (602.239).

В результате проверки гипотез авторы исследования выяснили:

1. Уровень вовлеченности в рекламный пост у крупных блогеров ниже, чем у мелких блогеров.
2. Средняя стоимость охвата 1000 человек в рекламном посте у крупных блогеров ниже, чем у мелких блогеров.
3. Уровень вовлеченности в рекламный пост у лайфстайл-блогеров и блогеров с узконаправленной тематикой не отличаются.
4. Средняя стоимость охвата 1000 человек в рекламном посте у лайфстайл-блогеров ниже, чем у блогеров с узконаправленной тематикой.
5. Средняя стоимость охвата 1000 человек в рекламном посте у блогеров с уровнем вовлеченности $\geq 2\%$ ниже, чем у блогеров с уровнем вовлеченности $< 2\%$.

На основе эмпирического исследования авторы утверждают, что при выборе блогеров нужно учитывать количество подписчиков и уровень вовлеченности в аккаунте. Если бренд хочет получить быстрый и дешевый охват в рекламной кампании, то авторы рекомендуют выбирать крупных блогеров с уровнем вовлеченности больше 2%.

Бренды не должны ориентироваться только на эти показатели. Лучше анализировать подход блогера к созданию контента, его личные качества, стилистику ведения аккаунта.

4 Заключение

На примере брендов рынка общественного питания Санкт-Петербурга авторы выяснили, что уровень вовлеченности пользователей в рекламный пост у крупных блогеров ниже. У крупных блогеров средняя стоимость охвата 1000 человек в рекламном посте ниже, чем у мелких блогеров. С ростом числа подписчиков стоимость охвата 1000 человек снижается.

Если бренд работает с массовым сегментом, то при выборе блогера можно менять одного крупного блогера на несколько мелких, чтобы получить более высокую вовлеченность. Такая замена подходит для показа рекламы на узкие (не массовые) сегменты, которые труднодоступны через крупных блогеров.

Для более демократичных и массовых сегментов нужны крупные блогеры. Широкая аудитория крупных блогеров позволяет донести информацию до большего числа людей с меньшими затратами.

Стоимость охвата 1000 человек у мелких блогеров выше, чем у крупных, но абсолютная стоимость размещения рекламы у мелких блогеров ниже. Если бренду нужна узкая аудитория, то лучше разместиться у мелкого блогера и заплатить больше за охват 1000 человек, но сэкономить на абсолютной стоимости размещения, т.к. в этом случае бренд не покупает широкую аудиторию крупного блогера. Если бренду нужно быстро охватить большую аудиторию и повысить узнаваемость, то лучше использовать крупных блогеров, т.к. они дадут больший охват дешевле, чем мелкие блогеры.

В дальнейшем исследовании авторы планируют включить в выборку бренды из других городов. Это позволит обобщить выводы для более широкого географического рынка. Авторы также планируют проверить влияние характеристик бренда (например, специализация на определенной кухне) на эффективность рекламы бренда у лидеров мнений.

Литература

1. Alexander E. Fashion Bloggers, Influencers, and Editors, Defined - What Is a Fashion Blogger? // *Harpers Bazaar* [Электронный ресурс]. URL: <https://www.harpersbazaar.com/uk/fashion/fashion-news/news/a41898/fashion-influencer-fashion-blogger-definition/> (дата обращения: 15.06.2020).
2. Arora A. [и др.]. Measuring social media influencer index- insights from facebook, Twitter and Instagram // *Journal of Retailing and Consumer Services*. 2019. (49). С. 86–101.
3. Brand Analytics Социальные сети в России: Цифры и тренды, осень 2018 // *Brand Analytics* [Электронный ресурс]. URL: <https://br-analytics.ru/blog/socseti-v-rossii-osen-2018/> (дата обращения: 14.09.2019).
4. Cooley D., Parks-Yancy R. The Effect of Social Media on Perceived Information Credibility and Decision Making // *Journal of Internet Commerce*. 2019. № 3 (18). С. 249–269.
5. eMarketer Instagram Influencer Share Worldwide, by Number of Followers, July 2018 (% of total) // *eMarketer* [Электронный ресурс]. URL: <https://www.emarketer.com/chart/226170/instagram-influencer-share-worldwide-by-number-of-followers-july-2018-of-total> (дата обращения: 14.09.2019).
6. eMarketer eMarketer Forecasts on Ad Blocking Usage Shows Stabilizing Growth // *eMarketer* [Электронный ресурс]. URL: <https://www.emarketer.com/content/ad-blocking-growth-is-slowing-down-but-not-going-away> (дата обращения: 16.09.2019).
7. eMarketer Who Do US Internet Users Trust for Recommendations When Shopping? // *eMarketer* [Электронный ресурс]. URL: <https://www.emarketer.com/chart/229941/who-do-us-internet-users-trust-recommendations-shopping-of-respondents-by-generation-feb-2019> (дата обращения: 16.09.2019).
8. IAB Coronavirus Ad Spend Impact: Buy-Side // *IAB* [Электронный ресурс]. URL: <https://www.iab.com/insights/coronavirus-ad-spend-impact-buy-side/> (дата обращения: 15.06.2020).
9. IAB RUSSIA Influence Marketing White Paper // *IAB RUSSIA* [Электронный ресурс]. URL: <https://iabrus.ru/projects/1157> (дата обращения: 16.09.2019).
10. Instagram Advertising on Instagram | Instagram for Business // *Instagram* [Электронный ресурс]. URL: <https://business.instagram.com/advertising/> (дата обращения: 18.09.2019).
11. Jin S.V., Muqaddam A. Product placement 2.0: “Do Brands Need Influencers, or Do Influencers Need Brands?” // *Journal of Brand Management*. 2019. № 5 (26). С. 522–537.
12. Katz E., Lazarsfeld P.F., Roper E. *Personal Influence* / E. Katz, P.F. Lazarsfeld, E. Roper, Routledge, 2017.
13. Ki C. ‘Chloe’, Kim Y. The mechanism by which social media influencers persuade consumers: The role of consumers’ desire to mimic // *Psychology & Marketing*. 2019. № 10 (36). С. 905–922.
14. Konstantopoulou A. [и др.]. Improving SMEs’ competitiveness with the use of Instagram influencer advertising and eWOM // *International Journal of Organizational Analysis*. 2019. № 2 (27). С. 308–321.
15. Livedune.ru Статистика социальных сетей в России 2020 // *Livedune.ru* [Электронный ресурс]. URL: https://livedune.ru/blog/statistika_socsetej_v_rossii (дата обращения: 07.07.2020).
16. Loureiro S.M.C., Sarmiento E.M. Exploring the Determinants of Instagram as a Social Network for Online Consumer-Brand Relationship // *Journal of Promotion Management*. 2019. № 3 (25). С. 354–366.
17. Martensen A., Brockenhuus-Schack S., Zahid A.L. How citizen influencers persuade their followers // *Journal of Fashion Marketing and Management: An International Journal*. 2018. № 3 (22). С. 335–353.
18. Merriam-Webster Definition of Blog by Merriam-Webster // *Merriam-Webster* [Электронный ресурс]. URL: <https://www.merriam-webster.com/dictionary/blog> (дата обращения: 19.09.2019).
19. Restoclub Рестораны Санкт-Петербурга на Restoclub.ru | Отзывы, обзоры, рейтинги ресторанов, кафе, баров, банкетных залов // *Restoclub* [Электронный ресурс]. URL: <https://www.restoclub.ru/> (дата обращения: 16.06.2020).
20. Rusability Тренды продвижения в Instagram Stories на 2019 год: масштабное исследование // *Rusability* [Электронный ресурс]. URL: <https://rusability.ru/internet->

- marketing/smm/trendy-prodvizheniya-v-instagram-stories-na-2019-god-masshtabnoe-issledovanie/ (дата обращения: 14.09.2019).
21. Rusability Пользователи соцсетей в России: статистика и портреты аудитории // Rusability [Электронный ресурс]. URL: <https://rusability.ru/internet-marketing/polzovateli-sotssetej-v-rossii-statistika-i-portrety-auditorii/> (дата обращения: 14.09.2019).
 22. Sostav.ru Из-за пандемии рекламодатели начали отдавать предпочтение рекламе у блогеров // Sostav.ru [Электронный ресурс]. URL: <https://www.sostav.ru/publication/iz-za-pandemii-reklamodateli-nachali-otdavay-predpochtenie-reklame-u-bloggerov-43361.html> (дата обращения: 07.07.2020).
 23. Tripadvisor Tripadvisor // Tripadvisor [Электронный ресурс]. URL: <https://www.tripadvisor.ru/> (дата обращения: 16.06.2020).
 24. Veirman M. De, Cauberghe V., Hudders L. Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude // International Journal of Advertising. 2017. № 5 (36). С. 798–828.
 25. Veirman M. De, Hudders L. Disclosing sponsored Instagram posts: the role of material connection with the brand and message-sidedness when disclosing covert advertising // International Journal of Advertising. 2019. С. 1–37.
 26. Welburn M. What Is Lifestyle Blogging? Plus Tips On Curating Your Own Unique Content «Blend» // theelginavenue [Электронный ресурс]. URL: <http://www.theelginavenue.com/thoughts-on-lifestyle-blogging-finding-your-blend-of-content.html> (дата обращения: 15.06.2020).
 27. Евсюкова Т.В., Гермашева Т.М. Основные подходы к определению понятий «блог» и «блогосфера» в лингвистике // Гуманитарные, социально-экономические и общественные науки. 2015. № 2 (11). С. 234–237.
 28. Иноземцева А. Почему стоит работать с микроблогерами, а не миллионниками. // Cossa.ru [Электронный ресурс]. URL: <https://www.cossa.ru/trends/216141/> (дата обращения: 18.09.2019).
 29. Лебедева В. Блогеры продадутся дороже. № 138 (6618) // Коммерсантъ. 2019.
 30. Мертон Р. Социальная теория и социальная структура / Р. Мертон, Москва: АСТ, 2006. 880 с.
 31. Мирошниченко М. Лидеры мнений: кто такие и как с ними работать // Cossa.ru [Электронный ресурс]. URL: <https://www.cossa.ru/trends/165308/> (дата обращения: 16.09.2019).
 32. Ткачук А. Зависимость ER в Instagram от числа подписчиков на выборке в 100 млн аккаунтов. Как изменился охват за последний месяц? // Dnative [Электронный ресурс]. URL: <https://dnative.ru/zavisimost-er-v-instagram-ot-chisla-podpischikov-na-vyborke-v-100-mln-akkauntov/> (дата обращения: 18.09.2019).
 33. Черний М. Блогеры и лидеры мнений в Instagram. Где искать, сколько платить, как анализировать // shcherbakovs.com [Электронный ресурс]. URL: <https://shcherbakovs.com/bloggers-and-opinion-leaders-in-instagram/> (дата обращения: 18.09.2019).

HOW OPINION LEADERS HELP BRANDS IN THE FOOD SERVICE MARKET

Alexandrovskiy Sergey Vladimirovich

Candidate of economic sciences

National Research University Higher School of Economics, assistant professor

Russian Federation, Moscow

saleksandrovskiy@hse.ru

Silkina Olga Sergeevna

National Research University Higher School of Economics, assistant professor

Russian Federation, Moscow

silkinaolgann@yandex.ru

Abstract

Instagram is one of the most popular online platforms among Russian users. When choosing products and services, consumers ask their friends for recommendations. Advertising posts with friendly recommendations from opinion leaders on Instagram have an impact on the purchase decision. Based on an empirical study of food service market in St. Petersburg, the authors found that the engagement rate in an advertising post is lower for big opinion leaders than for small ones; the average cost per thousand (ad reach) is lower for big opinion leaders, lifestyle bloggers, and opinion leaders with an engagement rate of more than 2%.

Keywords

Instagram, opinion leader, blogger, advertising, engagement rate, cost per thousand

References

1. Alexander E. Fashion Bloggers, Influencers, and Editors, Defined - What Is a Fashion Blogger? // Harpers Bazaar [Electronic resource]. URL: <https://www.harpersbazaar.com/uk/fashion/fashion-news/news/a41898/fashion-influencer-fashion-blogger-definition/> (accessed: 15.06.2020).
2. Arora A. [и др.]. Measuring social media influencer index- insights from facebook, Twitter and Instagram // Journal of Retailing and Consumer Services. 2019. (49). P. 86–101.
3. Brand Analytics Social'nye seti v Rossii: Cifry i trendy, osen' 2018 // Brand Analytics [Elektronnyj resurs]. URL: <https://br-analytics.ru/blog/socseti-v-rossii-osen-2018/> (accessed: 14.09.2019).
4. Cooley D., Parks-Yancy R. The Effect of Social Media on Perceived Information Credibility and Decision Making // Journal of Internet Commerce. 2019. № 3 (18). P. 249–269.
5. eMarketer Instagram Influencer Share Worldwide, by Number of Followers, July 2018 (% of total) // eMarketer [Electronic resource]. URL: <https://www.emarketer.com/chart/226170/instagram-influencer-share-worldwide-by-number-of-followers-july-2018-of-total> (accessed: 14.09.2019).
6. eMarketer eMarketer Forecasts on Ad Blocking Usage Shows Stabilizing Growth // eMarketer [Electronic resource]. URL: <https://www.emarketer.com/content/ad-blocking-growth-is-slowing-down-but-not-going-away> (accessed: 16.09.2019).
7. eMarketer Who Do US Internet Users Trust for Recommendations When Shopping? // eMarketer [Electronic resource]. URL: <https://www.emarketer.com/chart/229941/who-do-us-internet-users-trust-recommendations-shopping-of-respondents-by-generation-feb-2019> (accessed: 16.09.2019).
8. IAB Coronavirus Ad Spend Impact: Buy-Side // IAB [Electronic resource]. URL: <https://www.iab.com/insights/coronavirus-ad-spend-impact-buy-side/> (accessed: 15.06.2020).
9. IAB RUSSIA Influence Marketing White Paper // IAB RUSSIA [Electronic resource]. URL: <https://iabrus.ru/projects/1157> (accessed: 16.09.2019).
10. Instagram Advertising on Instagram | Instagram for Business // Instagram [Electronic resource]. URL: <https://business.instagram.com/advertising/> (accessed: 18.09.2019).

12. Katz E., Lazarsfeld P.F., Roper E. *Personal Influence* / E. Katz, P.F. Lazarsfeld, E. Roper, Routledge, 2017.
13. Ki P. 'Chloe', Kim Y. The mechanism by which social media influencers persuade consumers: The role of consumers' desire to mimic // *Psychology & Marketing*. 2019. № 10 (36). P. 905–922.
14. Konstantopoulou A. [etP.]. Improving SMEs' competitiveness with the use of Instagram influencer advertising and eWOM // *International Journal of Organizational Analysis*. 2019. № 2 (27). P. 308–321.
15. Livedune.ru Статистика социальных сетей в России 2020 // Livedune.ru [Electronic resource]. URL: https://livedune.ru/blog/statistika_socsetej_v_rossii (accessed: 07.07.2020).
16. Loureiro S.M.P., Sarmiento E.M. Exploring the Determinants of Instagram as a Social Network for Online Consumer-Brand Relationship // *Journal of Promotion Management*. 2019. № 3 (25). P. 354–366.
17. Martensen A., Brockenhuus-Schack S., Zahid A.L. How citizen influencers persuade their followers // *Journal of Fashion Marketing and Management: An International Journal*. 2018. № 3 (22). P. 335–353.
18. Merriam-Webster Definition of Blog by Merriam-Webster // Merriam-Webster [Electronic resource]. URL: <https://www.merriam-webster.com/dictionary/blog> (accessed: 19.09.2019).
19. Restoclub Restorany Sankt-Peterburga na Restoclub.ru | Otzyvy, obzory, rejtingi restoranov, kafe, barov, banketnyh zalov // Restoclub [Elektronnyj resurs]. URL: <https://www.restoclub.ru/> (accessed: 16.06.2020).
20. Rusability Trendy prodvizheniya v Instagram Stories na 2019 god: masshtabnoe issledovanie // Rusability [Elektronnyj resurs]. URL: <https://rusability.ru/internet-marketing/smm/trendy-prodvizheniya-v-instagram-stories-na-2019-god-masshtabnoe-issledovanie/> (accessed: 14.09.2019).
21. Rusability Pol'zovateli socsetej v Rossii: statistika i portrety auditorii // Rusability [Elektronnyj resurs]. URL: <https://rusability.ru/internet-marketing/polzovateli-sotssetej-v-rossii-statistika-i-portrety-auditorii/> (accessed: 14.09.2019).
22. Sostav.ru Iz-za pandemii reklamodатели nachali otdavat' predpochtenie reklame u blogerov // Sostav.ru [Elektronnyj resurs]. URL: <https://www.sostav.ru/publication/iz-za-pandemii-reklamodately-nachali-otdavayt-predpochtenie-reklame-u-blogerov-43361.html> (accessed: 07.07.2020).
23. Tripadvisor Tripadvisor // Tripadvisor [Electronic resource]. URL: <https://www.tripadvisor.ru/> (accessed: 16.06.2020).
24. Veirman M. De, Cauberghe V., Hudders L. Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude // *International Journal of Advertising*. 2017. № 5 (36). P. 798–828.
25. Veirman M. De, Hudders L. Disclosing sponsored Instagram posts: the role of material connection with the brand and message-sidedness when disclosing covert advertising // *International Journal of Advertising*. 2019. P. 1–37.
26. Welburn M. What Is Lifestyle Blogging? Plus Tips On Curating Your Own Unique Content «Blend» // theelginavenue [Electronic resource]. URL: <http://www.theelginavenue.com/thoughts-on-lifestyle-blogging-finding-your-blend-of-content.html> (accessed: 15.06.2020).
27. Evsyukova T.V., Germasheva T.M. Osnovnye podhody k opredeleniyu ponyatij «blog» i «blogosfera» v lingvistike // *Gumanitarnye, social'no-ekonomicheskie i obshchestvennye nauki*. 2015. № 2 (11). P. 234–237.
28. Inozemceva A. Pochemu stoit rabotat' s mikrobloggerami, a ne millionnikami. // Cossa.ru [Elektronnyj resurs]. URL: <https://www.cossa.ru/trends/216141/> (accessed: 18.09.2019).
29. Lebedeva V. Blogery prodadutsya dorozhe. № 138 (6618) // *Kommersant*". 2019.
30. Merton R. *Social'naya teoriya i social'naya struktura* / R. Merton, Moskva: AST, 2006. 880 P.
31. Miroshnichenko M. Lidery mnenij: kto takie i kak s nimi rabotat' // Cossa.ru [Elektronnyj resurs]. URL: <https://www.cossa.ru/trends/165308/> (accessed: 16.09.2019).
32. Tkachuk A. Zavisimost' ER v Instagram ot chisla podpischikov na vyborke v 100 mln akkauntov. Kak izmenilsya ohvat za poslednij mesyac? // Dnative [Elektronnyj resurs]. URL: <https://dnative.ru/zavisimost-er-v-instagram-ot-chisla-podpischikov-na-vyborke-v-100-mln-akkauntov/> (accessed: 18.09.2019).



33. Chernij M. Blogery i lidery mnenij v Instagram. Gde iskat', skol'ko platit', kak analizirovat' // shcherbakovs.com [Elektronnyj resurs]. URL: <https://shcherbakovs.com/bloggers-and-opinion-leaders-in-instagram/> (accessed: 18.09.2019).

Информационное общество и власть

ИНСТРУМЕНТЫ ЭЛЕКТРОННОГО УЧАСТИЯ ГРАЖДАН В ПУБЛИЧНОМ УПРАВЛЕНИИ: ПОНЯТИЕ И КЛАССИФИКАЦИЯ

Статья рекомендована к публикации членом редакционного совета М.В. Якушевым 30.09.2020.

Погодина Ирина Владимировна

Кандидат юридических наук, доцент

Федеральное государственное бюджетное образовательное учреждение высшего образования «Владимирский государственный университет им. А.Г. и Н.Г. Столетовых», Юридический институт, заведующая кафедрой «Финансовое право и таможенная деятельность»

Владимир, Россия

irinapogodina@mail.ru

Авдеев Данила Алексеевич

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Владимирский государственный университет им. А.Г. и Н.Г. Столетовых», Юридический институт, студент

Владимир, Россия

danilavdeev@yandex.ru

Аннотация

В статье авторы определяют понятие «инструменты электронного участия» и предлагают отвечающую потребностям науки и практики государственного управления классификацию данных инструментов.

Ключевые слова

Публичное управление, электронное участие, инструменты электронного участия

Введение

Несколько лет назад органы власти использовали Интернет преимущественно как площадку для одностороннего предоставления информации, любое интерактивное взаимодействие с пользователями сети практически исключалось. Сайты органов власти выполняли единственную функцию – информационного обеспечения их деятельности. Технический прогресс, а также рост ожиданий граждан по поводу их участия в публичном управлении обусловили расширение области использования информационно-коммуникационных технологий в административной сфере.

Современные технологии в административной деятельности применяются не только для дистанционного предоставления государственных (муниципальных) услуг, но и для принятия общественно-значимых решений, поддержки уже принятых, получения отзывов граждан, иных форм интерактивного взаимодействия власти и общества. Использование инструментов информационно-коммуникационных технологий в контексте участия общественности привело к появлению термина «электронное участие» [12; 13].

Участие политических субъектов, институтов гражданского общества и граждан в реализации изменений в системе государственного управления справедливо называется основным условием их устойчивости [5], тем более, что новая архитектура государственного управления должна быть основана на сотрудничестве и соуправлении [5]. В этом смысле информационно-коммуникационные технологии способны существенно повысить качество «обратной связи» и политического участия, делая их более оперативными, удобными, доступными, достоверными и экономичными [2].

© Погодина И.В., Авдеев Д.А., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

Необходимость теоретического осмысления вопросов электронного участия общественности в публичном управлении обусловлена также тем обстоятельством, что в настоящее время в нашей стране осуществляется большая работа по практическому внедрению различного рода электронных сервисов и технологий в государственное и муниципальное управление. Некоторые из инициатив направлены на создание сервисов электронного участия (в частности, реализация федерального проекта «Цифровое государственное управление» предусматривает создание информационных систем поддержки принятия решений высшими органами государственной власти).

Указанное свидетельствует о необходимости интенсификации научного поиска в сфере электронного участия граждан в публичном управлении. Несмотря на широкий спектр исследований в этой сфере (в рамках юриспруденции, экономики, политологии, государственного и муниципального управления), некоторые терминологические противоречия остаются. Они касаются, в частности, дефиниции и перечня инструментов электронного участия. Необходимость формулирования понятия связана с наличием нескольких подходов к данному термину в научных исследованиях. Иллюстрацией этому может служить то обстоятельство, что в некоторых работах не различаются инструменты электронного участия и смежные категории (формы, каналы, уровни электронного участия и т.д.), в других же публикациях авторы вовсе не используют данный термин при характеристике электронного участия.

Цель настоящей статьи – определить понятие «инструменты электронного участия» и предложить классификацию указанных инструментов, отвечающую потребностям науки и практики государственного управления.

1 Определение понятия «инструменты электронного участия»

Для уточнения смысла понятия «инструмент электронного участия» обратимся к словарной литературе. Историко-этимологический словарь латинских заимствований предлагает следующее определение слову «инструмент»: 1) орудие, преимущественно ручное, для производства каких-либо работ; 2) орудие, средство, преимущественно для достижения какой-либо цели [6]. Аналогичную трактовку предлагает Толковый словарь Ожегова [7]. Вполне очевидно, что применительно к сфере электронного участия «инструмент» следует трактовать во втором (переносном) значении. Учитывая, что платформой для реализации электронного участия граждан является Интернет [3], инструментами электронного участия выступают особые электронные сервисы (платформы, сайты, приложения и т.д.), позволяющие гражданам участвовать в публичном управлении дистанционно.

Принимая во внимание лексическое значение данного слова, едва ли правильно под «инструментами электронного участия» понимать различного рода способы, формы участия, каналы коммуникации между властью и обществом, что иногда наблюдается в научной литературе. Так, видится спорным определение инструментов электронного участия, предложенное М.В. Леоновой в диссертационном исследовании, согласно которому инструменты электронного участия и механизмы обратной связи определяются как такие способы взаимодействия пользователей с представителями власти, как методы получения мнений, методы участия граждан, методы контроля и т.д. [8]. На наш взгляд, в данной дефиниции смешиваются близкие, но все-таки разнопорядковые категории: инструменты, методы и способы электронного участия.

В некоторых исследованиях электронного управления авторы не используют термин «инструменты электронного участия», хотя описывают, на наш взгляд, именно их. К примеру, О.Н. Демушкина перечисляет сервисы и технологии, позволяющие «взаимодействовать с органами власти нашей страны и, в большей или меньшей степени, влиять на их действия», относя к ним различного рода сайты и платформы (страницы на сайтах органов власти, сервисы электронных петиций, краудсорсинговые платформы, технологии электронного голосования), а также блоги политиков и представителей органов власти [4]. Однако словосочетание «инструменты электронного участия» применительно к данным технологическим решениям автор не употребляет.

В зарубежной науке публичного управления изучение электронного участия имеет определённую традицию. К примеру, в исследовании DEMO-net¹ со ссылкой на работу А. Макинтош и соавторов приводится объёмный перечень инструментов электронного участия (24 позиции): начиная от блогов и механизмов оповещения до более сложных консультационных платформ [11]. При интерпретации данной работы следует учесть, что это исследование подготовлено почти 15 лет назад и технологии, безусловно, сделали шаг вперед. Некоторые из приведенных в работе инструментов ушли на второй план, появились и новые, не указанные в обзоре (например, мобильные приложения для смартфонов). Тем не менее, представленная палитра инструментов позволяет выявить сущность инструментов электронного участия. Видно, что общим признаком, который присущ данным инструментам, является цель их использования. Она соответствует цели электронного участия – «учет мнения граждан в государственном и муниципальном управлении, при принятии политических решений, включая реагирование на обращения граждан по широкому кругу проблем» [9].

На основании приведенных выше теоретических положений представляется возможным сформулировать следующую дефиницию инструментов электронного участия в публичном управлении. **Под ними следует понимать технологические решения (приложения для мобильных устройств, электронные сервисы, программы и т.п.), которые используются для участия граждан в государственном и муниципальном управлении с целью учёта мнения населения при принятии публичных решений.**

Инструменты электронного участия, в нашем понимании, не тождественны сервисам, позволяющим получить государственные (муниципальные) услуги в электронной форме, несмотря на то что развитие и первых, и вторых – результат цифровизации публичного управления. Между ними имеется существенная, прежде всего функциональная, разница: инструменты электронного участия предназначены для создания дополнительных форм (способов) взаимодействия правительств с заинтересованными лицами для обсуждения общественно-значимых вопросов, принятия соответствующих решений, и не нацелены на оказание гражданам индивидуальных услуг. Различие между двумя указанными типами Интернет-сервисов состоит также в том, что электронное участие может осуществляться на сторонних ресурсах, созданных третьими лицами (например, в социальных сетях), тогда как услуги электронного правительства, как правило, оказываются на специально созданных веб-сайтах, поддерживаемых органами власти и подконтрольных им².

2 Классификация инструментов электронного участия

Определив, что следует понимать под «инструментами электронного участия», перейдем к их классификации (рис.1). В зарубежной литературе предлагается классификация инструментов электронного участия, основанная на их функциональных свойствах, с учётом систем, типов приложений, лежащих в их основе [1]. Данный классификационный ряд – весьма длинный, так как исчерпывающий перечень привести вряд ли возможно, да и не целесообразно: с развитием технологий появляются новые инструменты, а некоторые, наоборот, перестают использоваться. В исследованиях последних пяти лет делается акцент на таких инструментах, как сообщества органов власти и местного самоуправления в социальных сетях, мобильных приложениях, электронных платформах общественного участия и некоторых других технологических решениях.

Исходя из такого классификационного основания, как цель создания электронного инструмента, можно выделить инструменты, специально созданные для электронного участия (например, платформы электронного общественного участия), и инструменты, адаптированные/используемые для этих целей соответствующими субъектами (например, социальные сети).

¹ DEMO-net (сокращенно от Democracy network) – это исследовательский проект сети научных учреждений, финансируемый шестой Рамочной программой Европейской комиссии. Цель DEMO-net.org – укрепление передового опыта научных, технологических и социальных исследований в области совместного управления путем интеграции исследовательского потенциала отдельных лиц и организаций, разбросанных по всей Европе. См.: [8].

² Между тем, существует практика (и её следует признать положительной) объединения двух типов интернет-площадок в рамках одной платформы, когда, к примеру, портал по оказанию публичных услуг предполагает возможность «обратной связи» от граждан.

С точки зрения закрепления в нормах права инструменты электронного участия можно разделить на следующие группы: предусмотренные действующим законодательством; не закрепленные в позитивном праве (но ему не противоречащие).

В зависимости от уровня власти, на котором осуществляется электронное участие, инструменты электронного участия делятся на муниципальные (местные), региональные (субъектов РФ), федеральные.

В зависимости от инициатора создания инструмента электронного участия их можно объединить в следующие группы: инструменты, созданные органами публичной власти (например, региональные и муниципальные платформы общественного участия); инструменты, созданные институтами гражданского общества и отдельными гражданами (к примеру, порталы петиций и т.д.). Причем в научных исследованиях отмечается, что «общество, как правило, создает проекты электронного взаимодействия раньше государства, эффективнее государства и более адресно» [1].

Также определённой прикладной значимостью обладает классификация инструментов электронного участия в зависимости от их назначения. В этом ряду можно выделить инструменты электронного участия, созданные для обсуждения нормативных правовых актов; получения органами власти отзывов о работе; мониторинга проблем; принятия решений.

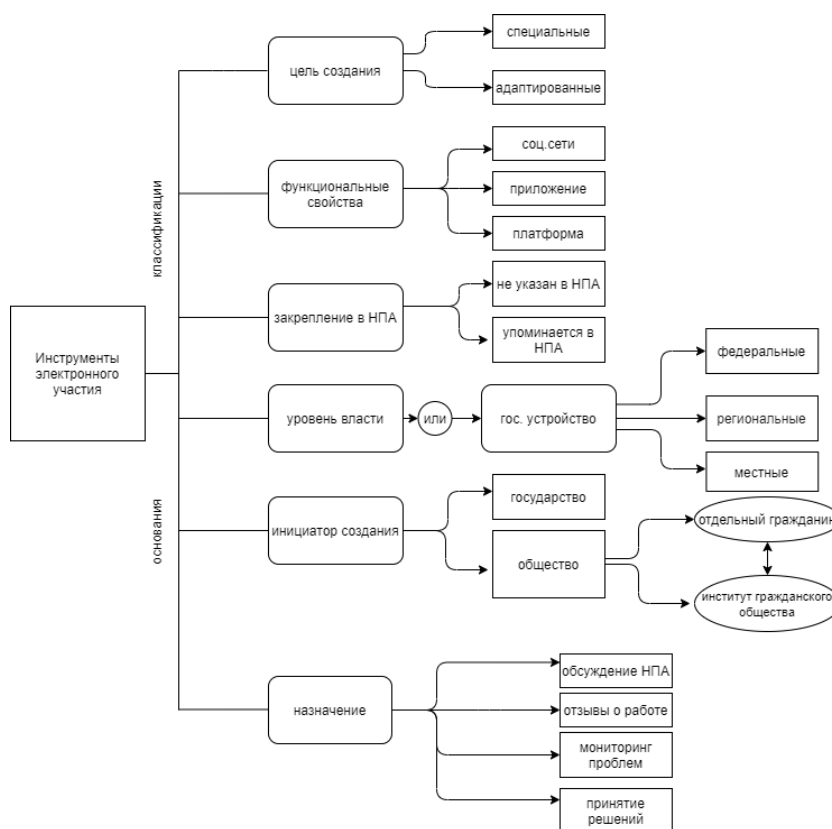


Рис. 1. Классификация инструментов электронного участия

Заключение

Таким образом, в настоящем исследовании мы сформулировали дефиницию инструментов электронного участия, а также предложили их классификацию по различным основаниям.

Вполне очевидно, что в условиях информатизации и цифровизации общества органы власти должны использовать инновационные инструменты для взаимодействия с гражданами. Есть основания полагать, что дальнейшая интеграция данных инструментов в административную

практику откроет новые возможности для трансформации отношений между гражданами, с одной стороны, и государственными органами и организациями, с другой. Информационно-коммуникационные технологии могут способствовать повышению вовлеченности граждан в публичное управление, создавая, таким образом, среду активного участия общественности в процессах принятия решений посредством электронных сервисов. В связи с этим теоретическая разработка вопросов электронного совместного публичного управления остаётся актуальным направлением научного поиска.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ и ЭИСИ в рамках научного проекта № 20-011-31680.

Литература

1. Волошинская А.А. «Электронное участие» в России: новый инструмент взаимодействия государства и общества или электронная потемкинская деревня? // Информационное общество. 2016. 1. С. 40-47.
2. Голубев С.В. «Электронное правительство»: проблемы эффективности политического управления (сравнительный анализ российского и мирового опыта): дисс. ... канд. полит. наук: 23.00.02. М., 225 с.
3. Голубева А.А., Ишматова Д.Р. Электронная демократия в России: формирование традиции политической осведомленности и участия // Вопросы государственного и муниципального управления. 2012. № 4. С.50-65.
4. Демушина О.Н. Анализ сервисов и платформ электронного участия граждан (на примере Волгоградской области) // Вопросы управления. 2017. №2 (45). С. 151-158.
5. Игнатова А.М. Политические и концептуальные основания трансформации государственного управления в современной России: дисс... канд. полит. наук: 23.00.02. Санкт-Петербург, 2019. С. 16.
6. Инструмент // Историко-этимологический словарь латинских заимствований. URL: <https://znachenie-slova.ru/инструмент> (дата обращения: 31.07.2020).
7. Инструмент // Толковый словарь Ожегова. URL: <https://dic.academic.ru/dic.nsf/ogegova/74770> (дата обращения: 30.07.2020).
8. Леонова М.В. Разработка методов и инструментов оценки качества механизмов электронного участия на сайтах электронного правительства: автореф. дисс... канд. техн. наук: 05.13.10. М., 2010. 22 с.
9. Чугунов А.В. Электронное участие в контексте трансформации концепции демократии // Управленческое консультирование. 2017. № 8 (104). С.14-20.
10. DEMO-net // Participedia. URL: <https://participedia.net/organization/316> (дата обращения: 30.07.2020).
11. DEMO-net. Introducing eParticipation. DEMO-net booklet series №1. 2007. URL: https://www.ifib.de/publikationsdateien/Introducing_eParticipation_DEMO-net_booklet_1.pdf (дата обращения: 30.07.2020).
12. Macintosh, A. Characterizing e-participation in policy-making // Proceedings of the 37th Annual Hawaii International Conference on System Sciences. 2004. Track 5. Vol. 5. P. 1-10.
13. Santamaría-Philco, A., Canós Cerdá, J. H., Penadés Gramaje, M. C. Advances in e-Participation: a perspective of Last Years // IEEE Access. Vol. 7. PP. 155894-155916. doi: 10.1109/ACCESS.2019.2948810.

TOOLS FOR THE ELECTRONIC PARTICIPATION OF CITIZENS IN PUBLIC GOVERNANCE: CONCEPT AND CLASSIFICATION

Pogodina Irina Vladimirovna

PhD in Law, associate professor

Vladimir State University named after Alexander and Nikolay Stoletovs, Law Institute, head of the Department of financial law and customs

Vladimir, Russia

irinapogodina@mail.ru

Avdeev Danila Alexeyevich

Vladimir State University named after Alexander and Nikolay Stoletovs, Law Institute, student

Vladimir, Russia

danilavdeev@yandex.ru

Abstract

In the article, the authors define the concept of “electronic participation instruments” and offer a classification of these tools that meets the needs of science and practice of public administration.

Keywords

Public government, e-participation, e-participation instruments

References

1. Voloshinskaya A.A. “Elektronnoye uchastiye” v Rossii: novyy instrument vzaimodeystviya gosudarstva i obshchestva ili elektronnoye potemkinskaya derevnya? // Informatsionnoye obshchestvo. 2016. 1. С. 40-47.
2. Golubev S.V. “Elektronnoye pravitel’stvo”: problemy effektivnosti politicheskogo upravleniya (sravnitel’nyy analiz rossiyskogo i mirovogo opyta): diss. ... kand. polit. nauk: 23.00.02. M., 225 s.
3. Golubeva A.A., Ishmatova D.R. Elektronnoye demokratiya v Rossii: formirovaniye traditsii politicheskoy osvedomlennosti i uchastiya // Voprosy gosudarstvennogo i munitsipal’nogo upravleniya. 2012. № 4. С. 50-65.
4. Demushina O.N. Analiz servisov i platform elektronnoye uchastiya grazhdan (na primere Volgogradskoy oblasti) // Voprosy upravleniya. 2017. №2 (45). S. 151-158.
5. Ignatova A.M. Politicheskiye i kontseptual’nyye osnovaniya transformatsii gosudarstvennogo upravleniya v sovremennoy Rossii: diss... kand. polit. nauk: 23.00.02. Sankt-Peterburg, 2019. S. 16.
6. Instrument // Istoriko-etimologicheskoy slovar’ latinskikh zaimstvovaniy. URL: <https://znachenie-slova.ru/instrument> (data obrashcheniya: 31.07.2020).
7. Instrument // Tolkovyy slovar’ Ozhegova. URL: <https://dic.academic.ru/dic.nsf/ogegova/74770> (data obrashcheniya: 30.07.2020).
8. Leonova M.V. Razrabotka metodov i instrumentov otsenki kachestva mekhanizmov elektronnoye uchastiya na saytakh elektronnoye pravitel’stva: avtoref. diss... kand. tekhn. nauk: 05.13.10. M., 2010. 22 s.
9. Chugunov A.V. Elektronnoye uchastiye v kontekste transformatsii kontseptsii demokratii // Upravlencheskoye konsul’tirovaniye. 2017. № 8 (104). S. 14-20.
10. DEMO-net // Participedia. URL: <https://participedia.net/organization/316> (date accessed: 07/30/2020).
11. DEMO-net. Introducing eParticipation. DEMO-net booklet series # 1. 2007. URL: https://www.ifib.de/publikationsdateien/Introducing_eParticipation_DEMO-net_booklet_1.pdf (date accessed: 30.07.2020).
12. Macintosh, A. Characterizing e-participation in policy-making // Proceedings of the 37th Annual Hawaii International Conference on System Sciences. 2004. Track 5. Vol. 5.P. 1-10.
13. Santamaría-Philco, A., Canós Cerdá, J. H., Penadés Gramaje, M. C. Advances in e-Participation: a perspective of Last Years // IEEE Access. Vol. 7. PP. 155894-155916. doi: 10.1109 / ACCESS.2019.2948810.

Образование в информационном обществе

ОБУЧЕНИЕ ДАТА-ГРАМОТНОСТИ: СОДЕРЖАНИЕ И КОНКУРЕНТНЫЕ ПРЕИМУЩЕСТВА ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Статья рекомендована к публикации членом редакционного совета А.М. Елизаровым 15.12.2020.

Дерябин Андрей Александрович

MSc Social Psychology

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Федеральный институт развития образования, Научно-исследовательский центр социализации и персонализации образования детей, научный сотрудник

Москва, Россия

deryabin-aa@ranepa.ru

Попов Александр Анатольевич

Доктор философских наук, доцент

Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Федеральный институт развития образования, Научно-исследовательский центр социализации и персонализации образования детей, заведующий научно-исследовательским сектором «Открытое образование»

Московский городской педагогический университет, Институт системных проектов, заведующий лабораторией компетентностных практик образования

Новосибирский государственный технический университет, гуманитарный факультет, профессор кафедры социологии и массовых коммуникаций

Москва, Россия

popov-aa@ranepa.ru

Аннотация

Представлен анализ зарубежного опыта и основных дискуссионных тем в разработке образовательных модулей науки о данных и машинного обучения для подросткового и юношеского возраста. Особое внимание уделяется таким практическим аспектам разработки содержания обучения как использование аутентичных данных в процессе обучения и обеспечение релевантности данных интересам учащихся. Приводятся характеристики учебных данных с точки зрения их сложности и эффекта на дата-грамотность, критическое мышление и предметные знания учеников. Рассматриваются условия становления субъектности учащихся через культивирование активной позиции «дата-продюсера» в противоположность традиционным дисциплинарным образовательным подходам. Материалом для анализа послужили преимущественно публикации 2014-2020 годов в научных изданиях, индексируемых в базе данных Scopus. Статья подготовлена в рамках выполнения научно-исследовательской работы государственного задания РАНХиГС.

Ключевые слова

дата-грамотность; наука о данных; машинное обучение; образовательные программы; дополнительное образование; анализ данных; data science; data literacy; искусственный интеллект; профориентация; информатика; обществознание

Введение

В контексте фундаментальных изменений в экономике и на рынке труда важное значение приобретает внедрение образовательных программ в области анализа данных и машинного обучения. Поддержание конкурентоспособности экономики требует широкого внедрения технологий искусственного интеллекта и обеспечение этой сферы соответствующими кадрами. Сформировать ответ на вызов сохранения конкурентоспособности призван ряд правительственных инициатив в научно-технологической и образовательной политике, в частности, «Национальная стратегия развития искусственного интеллекта на период до 2030 года». Основными

© Дерябин А.А., Попов А.А., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

направлениями повышения уровня обеспечения российского рынка технологий искусственного интеллекта квалифицированными кадрами в ней называются: (а) разработка и внедрение образовательных модулей в рамках образовательных программ всех уровней образования для получения гражданами знаний, приобретения ими компетенций и навыков в области математики, программирования, анализа данных, машинного обучения, способствующих развитию искусственного интеллекта. При этом в целях развития перспективных методов искусственного интеллекта приоритетное значение приобретает конвергентное знание, обеспечиваемое в том числе за счет интеграции математического, естественно-научного и социально-гуманитарного образования.¹ В соответствии с этой повесткой и с учетом динамично развивающейся индустрии данных российские образовательные программы требуют содержательного обновления.

В российском ландшафте школьных дисциплин дата-грамотность и наука о данных как практика аналитического и количественного мышления не имеют пока собственного места. Математика в школе, сосредоточившись на базовой теории, сопротивляется освоению практических методов ее применения в реальной жизни. В школьной физике и биологии данным отводится иллюстративная роль несмотря на то, что это науки, в которых важную роль играет эксперимент. Общественные дисциплины, всё более дата-ориентированные на практике, едва ли сегодня готовы к тому, чтобы передать школе осмысленное применение данных на уроках обществознания. Пока образовательное сообщество решает задачу, как нужно изменить образовательные системы, чтобы из школы вышли ученики с концептуальным пониманием и навыками анализа данных, наиболее эффективной формой для этого остается дополнительное образование.

Сфера дополнительного образования, в свою очередь, является конкурентной средой, в которой важным критерием востребованности образовательных программ является их содержание и качество. Понятие «дата-грамотность», спектр компетенций в области науки о данных и примеры существующих образовательных программ для учащихся подросткового и юношеского возраста обсуждались нами ранее.² Настоящая статья посвящена тому, как важнейшие аспекты содержания образовательной программы по дата-грамотности, а именно, характер учебных данных и тематика учебных проектов делают обучение более эффективным, придавая образовательному продукту конкурентное преимущество на рынке образовательных услуг.

1 Какие данные использовать для обучения?

С точки зрения ряда специалистов, наибольший учебный эффект на дата-грамотность, критическое мышление и предметные знания старшеклассников имеет погружение в исследование аутентичных научных данных [1]. Авторы широко цитируемого подхода «дата-шагов» Erickson и др. считают, что вводный курс науки о данных должен давать учащимся задачи с разнообразными по содержанию, необработанными, «непослушными» данными [2]. Wolff и др. [49], работая с детьми 10-14 лет, также полагают, что следует предъявлять учащимся не упрощенные, но комплексные данные. Однако «комплексные» данные не обязательно означает «большие», а «реальные» не обязательно означает «подлинные». Чтобы избежать путаницы в терминах, приводим ниже адресованный преподавателям разбор характеристик аутентичных учебных данных, обобщенный в Таблице 1.

Под аутентичными, или подлинными, данными понимаются «достоверные количественные и качественные данные, извлеченные из феноменов реальной жизни». Наоборот, неаутентичные данные могут быть искусственно сгенерированы в демонстрационных целях или могут быть результатом манипулирования данными для достижения задуманного учебного результата или интерпретации. Аутентичные данные могут собраны с использованием различных измерительных методов и инструментов (например, данные сенсоров), сгенерированы компьютерными моделями и симуляциями или получены из онлайн-репозиторий и научных публикаций [4, 5, 6]. Kjelvik и Schultheis [1] имеют в виду в первую очередь естественнонаучные данные, однако с некоторыми допущениями и в зависимости от содержания учебной программы преподаватель вполне может отнести к ним неадаптированные экономические, социологические и проч. данные. Сбор, анализ,

¹ Указ Президента Российской Федерации от 10.10.2019 г. № 490 О развитии искусственного интеллекта в Российской Федерации [Электронный ресурс] / Сайт Президента РФ. URL: <http://www.kremlin.ru/acts/bank/44731> (дата обращения 10.04.2020).

² Дерябин А.А., Попов А.А. Дата-грамотность как новая цифровая компетенция // Информационное общество, 2020, № 5. С. 39-47.

интерпретация аутентичных датасетов потенциально помогают учащимся улучшить навыки критического мышления, углубить предметные знания и связать свои количественные аналитические навыки со сферой науки о данных [7, 8].³ Ниже представлены характеристики сложности учебных данных, которые могут варьироваться в зависимости от задач учебной программы и подготовленности аудитории.

Охват (scope) данных определяется числом переменных и числом значений, которые принимает каждая переменная. Простые датасеты состоят из небольшого числа переменных и содержат только ту информацию, которая нужна для ответа на вопросы исследования. Сложные, с широким охватом датасеты могут содержать много переменных, как нужных, так и не нужных для решения поставленной исследовательской задачи. «Узкие» датасеты хороши для наглядной демонстрации зависимых и независимых переменных и отношений между ними; работа с «широкими» датасетами может представлять собой дополнительный вызов для учащихся, которые должны определить, какие переменные им нужны, а какие нет, и может стимулировать их исследовательскую активность, которая может вести к творческой переформулировке поставленных задач и выявлению непредвиденных связей между переменными.

Селективность (selection) касается принятия решений о том, какие переменные нужно определить, какие данные, откуда и при помощи каких инструментов необходимо собрать для решения поставленной исследовательской задачи. В простейшем случае переменные задаются преподавателем. В случае, если учащимся предлагается сделать это самостоятельно, им, тем самым, предоставляется возможность самим определить свою образовательную траекторию [9]. Последнее, разумеется, обеспечивает более богатый образовательный и исследовательский опыт [10, 8, 11].

Курирование (curation) данных – это организация и интеграция данных, собранных из разных источников, подготовка их к анализу и обеспечение их целостности в этом процессе; курирование может включать в себя обнаружение и сбор данных, аннотирование, подготовку и очистку, представление, организацию доступа к данным и их хранение. Преподаватель зачастую обеспечивает учащихся хорошо организованными, готовыми к использованию датасетами, чтобы облегчить их задачу. Однако, когда курирование делегируется студентам, у них появляется возможность освоить новые навыки работы с данными [10]: синтез нескольких датасетов в один предполагает приведение их к единому формату и размерности, кодирования категориальных переменных в численные и проч., что требует умений работы с электронными таблицами и соответствующими программными методами.

Беспорядок (messiness) – наличие в датасете пропусков и выбросов, случайных и систематических ошибок, влияющих на изменчивость переменных. Подготовка сырых данных – обычно едва ли не самый длительный этап в цикле промышленного дата-проекта, и навыки чистки «грязных» данных имеют большую ценность как для лучшего понимания учащимися источников изменчивости признаков в датасете, так и для приобретения практических навыков манипулирования данными.

Размер (size) данных определяется числом переменных и числом значений переменной (например, в случае табличных данных, это число столбцов и строк соответственно). В учебном процессе размер учебного датасета может варьироваться от маленького на начальных этапах обучения, до большого в ходе выполнения студенческого проекта.

Обеспечение учащихся опытом работы с разными типами данных вместе с обсуждением контекста использования этих данных и того, как именно эти данные помогают ответить на вопросы их исследования, может быть важным для развития дата-грамотности [12]. Ряд педагогических инициатив позволяет предположить, что взаимодействие с аутентичными данными положительно влияет на дата-грамотность учащихся [9, 13]. Вместе с тем, недостаток исследований методик, использующих аутентичные данные, не позволяет сделать вывод об однозначной эффективности этого подхода [14]. Исследований, сравнивающих эффекты аутентичных и неаутентичных данных на результаты обучения в настоящее время нами не обнаружено. Чтобы продемонстрировать эту зависимость, требуются исследования, которые бы изолировали воздействие аутентичных данных на дата-грамотность, интересы учащихся и их вовлеченность в изучение данных через манипулирование использованием аутентичных и неаутентичных данных в процессе обучения.

³ Цит. по [1].

Таблица 1. Характеристики сложности учебных данных по Kjelvik и Schultheis [1].

Характеристика	← Простота		Сложность →	
Охват (scope)	Узкий: есть только нужные данные		Широкий: есть нужные и ненужные данные	
Селективность (selection)	Переменные заданы преподавателем	Студенты выбирают переменные из предложенного датасета	Студенты самостоятельно собирают датасет с нужными им переменными	
Курирование (curation)	Полное: датасет дается студентам в готовом к анализу виде	Частичное: сырые данные, готовые к анализу, но не аннотированные	Отсутствует: студенты должны аннотировать сырые данные и подготовить их к анализу, трансформировать	Синтез: студенты должны объединить несколько датасетов и подготовить их к анализу
Размер (size)	Маленький: несколько переменных и их значений		Большой: много переменных и их значений	
Беспорядок (messiness)	Чистые: нет пропусков данных и выбросов, низкая изменчивость		Грязные: есть пустые значения и выбросы, высокая изменчивость	

2 Релевантность данных интересам учащихся как условие эффективной реализации образовательной задачи

Для эффективного обучения данные и результаты работы с ними должны быть «присвоены» студентами, а для этого они должны быть не только аутентичными, но и релевантными интересам учащихся. Когда учащиеся не представляют, как данные, с которыми им предложили работать, связаны с реальной жизнью, учебные задачи и действия могут представляться им затруднительными или бессмысленными [1]. Некоторые могут испытывать трудности с интерпретацией неаутентичных данных [15, 16] или с применением новых умений в других контекстах [17].

Напротив, аутентичные данные, связанные с предметным содержанием, с понятным контекстом их создания и в ясной связи с проблемами реального мира потенциально интересней и привлекательней для учащихся [12, 18], и от них поступает обратная связь, которая говорит об эмоциональной вовлеченности в работу с такими данными [8, 19].

Как обеспечить это присвоение данных в образовательном процессе? Через релевантную интересам учеников тематику учебных задач, сопутствующих им данных и взаимодействие с ними. Образовательная программа должна предусматривать решение таких образовательных задач, которые, с одной стороны, имеют статус «настоящих» (то есть актуальных и практически значимых) для всех ее участников, а с другой стороны, задают некоторую неопределенность как в ее интерпретации, так и в сценариях их решения.⁴ Это может быть привнесено в образовательный процесс как за счет самостоятельного определения учащимися целей и задач их дата-проекта, так и за счет использования необработанных аутентичных данных. Такая специфика программы позволяет обеспечить продуктивную деятельность учащихся и удерживать их высокую мотивацию на использование данных и техник работы с данными как с образовательным ресурсом.

Wilkerson и Polman [20] отмечают, что изучение науки о данных – это не просто утилитарный процесс предметного обучения, но выстраивание многослойных отношений учащегося по отношению к данным как социокультурному тексту: каждый учащийся может обнаружить данные, которые рассказывают о нем что-то важное, будь то данные от сенсоров носимых им «умных часов», геометки в истории его навигатора или обобщенная социально-экономическая статистика о регионе его проживания.

Lee и Dubovi [21] показывают, как возникает глубокая и сложная связь с данными у семей с диабетом 1 типа в процессе использования ими устройств мониторинга и постоянных измерений.

⁴ Глухов П.П. Компетентностные испытания как современная форма оценки образовательных достижений // Философия образования, 2016, № 4 (67). С. 99-110.

Kahn [22] описывает конструирование «семейных геобиографий» на основе данных и демонстрирует, как их интерпретации в сочетании с семейными историями миграции формируют вокруг них многослойный социоисторический контекст, влияющий на молодое поколение.

Кроме того, для создания условий эффективной профессиональной пробы важно дать учащимся попробовать себя в разных позициях в ходе решения задач по отношению к данным – например, в разных ролях в соответствии с разделением труда внутри дата-проекта (программист, дата-инженер, менеджер проекта и предметный эксперт). Так, Wise [23], описывая свой проект, в котором студенты собирают данные о собственном обучении, выделяет три взаимодополняющих позиции по отношению к данным, между которыми переключаются ее студенты в ходе учебного проекта: «внутренняя» (респондент, источник данных), «внешняя» (реципиенты, которым адресован аналитический отчет о проделанной работе, например «заказчик») и «аналитик данных» (тот, кто работает с данными).

Hardy [24] подчеркивает важность становления субъектности учащихся в генерации знания через культивирование активной позиции «дата-продюсера» в противоположность традиционным, дисциплинарным образовательным подходам, нацеленным на воспроизводство заранее известного и дидактически «правильного» результата и «правильных» данных, которые словно бы ждут, чтобы их собрали.

3 Метапредметные компетенции: «думать данными»

Говоря о профориентационной компоненте преподавания науки о данных в подростковом и юношеском возрасте, следует учитывать актуальные требования, предъявляемые к современному специалисту: не только соотносить конкретную задачу с перечнем испытанных вариантов решения и выбирать из этого перечня оптимальный вариант, но и конструировать новые способы решения, оптимальные именно для данного конкретного случая; оформлять проблемную ситуацию и ставить задачи, исходя из неё; намечать необходимый результат для своих действий в рамках конкретного проекта или технологического процесса, и выстраивать программу действий по достижению этого результата. Для этого у специалиста должен быть сформирован тип мышления, соответствующий выбранной им сфере профессиональной деятельности, позволяющий понимать присущие этой сфере технологии, приёмы работы и тип результата в их взаимосвязи, как динамическую систему, способную меняться и развиваться как вследствие обновления технологий, так и вследствие изменения требований к конечному продукту.⁵

Так, Суосо [25] пишет о важности формирования «ментальной привычки» как способа мышления, вопрошания, и решения задач. Finzer [26] также приверженец идеи, что в обучении в школе нужно прививать специфическое мышление, взгляд на мир сквозь призму данных. Это видение прослеживается и у авторов, пишущих о разработке образовательных программ по Data Science и статистике на уровне высшей школы [27, 28].

Заключение

В современных формах инженерного образования принципиальным становится требование к студенту: для того, чтобы профессиональное знание было эффективно освоено и присвоено, уже на старте получения профессионального образования студент должен иметь образ будущей профессии, понимать, каким образом осваиваемые знания и способы усиливают его как будущего специалиста. Следовательно, работа по формированию у будущих дата-аналитиков специфически профессионального мышления и основ профессиональной культуры должна происходить ещё в период их обучения в школе, в режиме пропедевтики будущей деятельности.

Формирование образа будущей профессии дата-аналитика и связанного с ней типа мышления может обеспечиваться практико-ориентированным образованием, где в качестве единицы при построении учебного процесса используется реально существующая проблема, которая переживается учеником как значимая.⁶ Эта проблема может быть описана посредством аутентичных количественных данных и оформлена в виде образовательной задачи, организующей

⁵ Попов А.А. Образовательные программы и элективные курсы компетентностного подхода / Изд. 6-е. – М.: Ленанд, 2019. – С.170.

⁶ Глухов П.П. Практическое мышление как компетентностный образовательный результат / Социальные коммуникации и эволюция обществ. Материалы V Международной научно-практической конференции. Под редакцией И.А. Вальдмана; Новосибирск: 2016. С. 447-453.

одновременно пробное практическое действие ученика и его рефлексии в отношении используемых им инструментов дата-аналитики.

Литература

1. Kjelvik M. K., Schultheis, E. H. Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. // *CBE Life Sciences Education*, 2019, 18(2). С. 1–8.
2. Erickson T. и др. Data Moves: one key to data science at school level / *Proceedings of the International Conference on Teaching Statistics (ICOTS-10)*, 2018. [Электронный ресурс]. URL https://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018 (дата обращения 10.04.2020).
3. 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, 129(March). С. 41–54. doi:10.1016/j.ijhcs.2019.03.006
4. Magnusson S. J. и др. How should learning be structured in inquiry-based science instruction? Investigating the interplay of 1st- and 2nd-hand investigations / In *Proceedings of the 6th international conference on learning sciences*. Alpharetta, GA: International Society of the Learning Sciences, 2004. С. 318–325.
5. Hug B., McNeill K. L. Use of first-hand and second-hand data in science: Does data type influence classroom conversations? // *International Journal of Science Education*, 2008, 30. С. 1725–1751.
6. Kerlin, S. C., McDonald, S. P., & Kelly, G. J. Complexity of secondary scientific data sources and students' argumentative discourse // *International Journal of Science Education*, 2010, 32(9). С. 1207–1225.
7. Mourad T., Grant B. W., Gram W. K. Engaging undergraduate students in ecological investigations using large, public datasets // *Teaching Issues and Experiments in Ecology*, 2012, 8.
8. Langen T. A. и др. Using large public datasets in the undergraduate ecology classroom // *Frontiers in Ecology and the Environment*, 2014, 12. С. 362–363.
9. Gould R., Sunbury S., Dussault M. (2014). In praise of messy data // *Science Teacher*, 2014, 81(8). С. 31.
10. Carlson J. R. и др. Determining data information literacy needs: A study of students and research faculty. [Электронный ресурс]. // *Libraries Faculty and Staff Scholarship and Research*, Paper 23. URL https://docs.lib.purdue.edu/lib_fsdocs/23/ (дата обращения 10.04.2020)
11. Calzada Prado, J., Marzal, M. Incorporating data literacy into information literacy programs: Core competencies and contents // *Libri*, 2013, 63(2). С. 123–134.
12. Schultheis E. H., Kjelvik, M. K. Data Nuggets: Bringing real data into the classroom to unearth students' quantitative and inquiry skills // *American Biology Teacher*, 2015, 77(1). С. 19–29.
13. Kastens K. A., Krumhansel R., Baker I. Thinking big – Transitioning your students from working with small, student-collected data sets towards “big data.” // *Science Teacher*, 2015, 82(5). С. 25–31.
14. Aikens M. L., Dolan E. L. Teaching quantitative biology: Goals, assessments, and resources // *Molecular Biology of the Cell*, 2014, 25. С. 3478–3481.
15. Piatek-Jimenez K. и др. Helping students become quantitatively literate. / *Mathematics Teacher*, 105(9), 2012. С. 692–696.
16. Common Core State Standards for Mathematics [Электронный ресурс]. // *Common Core State Standards Initiative (CCSSI)*, 2014. URL <http://www.corestandards.org/Math> (дата обращения 10.04.2020)
17. Borges-Rey E. L. Data literacy and citizenship: Understanding “big data” to boost teaching and learning in science and mathematics / Ramírez-Montoya M.-S. (Ed.), *Handbook of research on driving STEM learning with educational technologies*. Hershey, PA: IGI Global, 2017. - С. 65–79.
18. Hulleman C. S., Harackiewicz J. M. Promoting interest and performance in high school science classes // *Science*, 2009, 326(5958). С. 1410–1412.
19. Wolff и др. Creating an Understanding of Data Literacy for a Data-driven Society [Электронный ресурс]. // *The Journal of Community Informatics*, 2016, 12(3). URL <http://ci-journal.net/index.php/ciej/article/view/1286> (дата обращения 10.04.2020).

20. Wilkerson M. H., Polman J. L. Situating Data Science: Exploring How Relationships to Data Shape Learning. *Journal of the Learning Sciences*, 2020, 29(1). С. 1–10.
doi:10.1080/10508406.2019.1705664
21. Lee V. R., Dubovi I. At home with data: Family engagements with data involved in type 1 diabetes management // *Journal of the Learning Sciences*, 2020, 29(1). С. 11-31.
doi:10.1080/10508406.2019.1666011
22. Kahn J. Learning at the Intersection of Self and Society: The Family Geobiography as a Context for Data Science Education // *Journal of the Learning Sciences*, 2020, 29 (1), С. 57-80.
23. Wise A. F. Educating Data Scientists and Data Literate Citizens for a New Generation of Data // *Journal of the Learning Sciences*, 2020, 29(1). С. 165–181.
24. Hardy L., Dixon C., Hsi S. From Data Collectors to Data Producers: Shifting Students' Relationship to Data // *Journal of the Learning Sciences*, 2020, 29(1). С. 104–126.
25. Cuoco A., Goldenberg E. P., Mark J. Habits of Mind: an organizing principle for mathematics curriculum // *Journal of Mathematical Behavior*, 1997, 15(4). С. 375-402.
26. Finzer, W. The data science education dilemma [Электронный ресурс]. // *Technology Innovations in Statistics Education*, 2013, 7(2). URL <https://escholarship.org/uc/item/7gv0q9dc> (дата обращения 10.04.2020).
27. De Veaux, R. и др. Curriculum guidelines for undergraduate programs in data science // *Annual Review of Statistics and Its Application*, 2017, 4(1). С. 15-30.
28. Hardin J. и др. Data science in statistics curricula: Preparing students to “think with data” // *The American Statistician*, 2015, vol. 69(4). С. 343-353.

TEACHING DATA LITERACY: CURRICULUM CONTENT AND COMPETITIVE ADVANTAGES

Deryabin, Andrey Aleksandrovich

MSc Social Psychology

Russian Presidential Academy of National Economy and Public Administration, Federal Education Development Institute, researcher

Moscow, Russia

deryabin-aa@ranepa.ru

Popov, Aleksandr Anatolyevich

Dr. Sci. (Philos.), associate professor

Russian Presidential Academy of National Economy and Public Administration, Federal Education Development Institute, head of open education section

Moscow City University, Institute of system projects, head of competency practices lab

Novosibirsk State Technical University, sociology & mass communications dept., professor

Moscow, Russia

popov-aa@ranepa.ru

Abstract

A review of publications and an analysis of major discussions about the Data Science curriculum development for middle and secondary school is presented with special emphasis on use of authentic data and data relevancy to students' context.

Keywords

big data; data literacy; data science education; machine learning; artificial intelligence; school education; computer science education; curriculum development

References

1. Kjelvik M. K., Schultheis, E. H. Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. // *CBE Life Sciences Education*, 2019, 18(2). С. 1–8.
2. Erickson T. и др. Data Moves: one key to data science at school level / *Proceedings of the International Conference on Teaching Statistics (ICOTS-10)*, 2018. [Электронный ресурс]. URL https://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018 (дата обращения 10.04.2020).
3. 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, 129(March). С. 41–54. doi:10.1016/j.ijhcs.2019.03.006
4. Magnusson S. J. и др. How should learning be structured in inquiry-based science instruction? Investigating the interplay of 1st- and 2nd-hand investigations / *In Proceedings of the 6th international conference on learning sciences*. Alpharetta, GA: International Society of the Learning Sciences, 2004. С. 318–325.
5. Hug B., McNeill K. L. Use of first-hand and second-hand data in science: Does data type influence classroom conversations? // *International Journal of Science Education*, 2008, 30. С. 1725–1751.
6. Kerlin, S. C., McDonald, S. P., & Kelly, G. J. Complexity of secondary scientific data sources and students' argumentative discourse // *International Journal of Science Education*, 2010, 32(9). С. 1207–1225.
7. Mourad T., Grant B. W., Gram W. K. Engaging undergraduate students in ecological investigations using large, public datasets // *Teaching Issues and Experiments in Ecology*, 2012, 8.
8. Langen T. A. и др. Using large public datasets in the undergraduate ecology classroom // *Frontiers in Ecology and the Environment*, 2014, 12. С. 362–363.

9. Gould R., Sunbury S., Dussault M. (2014). In praise of messy data // *Science Teacher*, 2014, 81(8). С. 31.
10. Carlson J. R. и др. Determining data information literacy needs: A study of students and research faculty. [Электронный ресурс]. // *Libraries Faculty and Staff Scholarship and Research, Paper 23*. URL https://docs.lib.purdue.edu/lib_fsdocs/23/ (дата обращения 10.04.2020)
11. Calzada Prado, J., Marzal, M. Incorporating data literacy into information literacy programs: Core competencies and contents // *Libri*, 2013, 63(2). С. 123–134.
12. Schultheis E. H., Kjellvik, M. K. Data Nuggets: Bringing real data into the classroom to unearth students' quantitative and inquiry skills // *American Biology Teacher*, 2015, 77(1). С. 19–29.
13. Kastens K. A., Krumhansel R., Baker I. Thinking big – Transitioning your students from working with small, student-collected data sets towards “big data.” // *Science Teacher*, 2015, 82(5). С. 25–31.
14. Aikens M. L., Dolan E. L. Teaching quantitative biology: Goals, assessments, and resources // *Molecular Biology of the Cell*, 2014, 25. С. 3478–3481.
15. Piatek-Jimenez K. и др. Helping students become quantitatively literate. / *Mathematics Teacher*, 105(9), 2012. С. 692–696.
16. Common Core State Standards for Mathematics [Электронный ресурс]. // *Common Core State Standards Initiative (CCSSI)*, 2014. URL <http://www.corestandards.org/Math> (дата обращения 10.04.2020)
17. Borges-Rey E. L. Data literacy and citizenship: Understanding “big data” to boost teaching and learning in science and mathematics / Ramírez-Montoya M.-S. (Ed.), *Handbook of research on driving STEM learning with educational technologies*. Hershey, PA: IGI Global, 2017. - С. 65–79.
18. Hulleman C. S., Harackiewicz J. M. Promoting interest and performance in high school science classes // *Science*, 2009, 326(5958). С. 1410–1412.
19. Wolff и др. Creating an Understanding of Data Literacy for a Data-driven Society [Электронный ресурс]. // *The Journal of Community Informatics*, 2016, 12(3). URL <http://ci-journal.net/index.php/ciej/article/view/1286> (дата обращения 10.04.2020).
20. Wilkerson M. H., Polman J. L. Situating Data Science: Exploring How Relationships to Data Shape Learning. *Journal of the Learning Sciences*, 2020, 29(1). С. 1–10. doi:10.1080/10508406.2019.1705664
21. Lee V. R., Dubovi I. At home with data: Family engagements with data involved in type 1 diabetes management // *Journal of the Learning Sciences*, 2020, 29(1). С. 11–31. doi:10.1080/10508406.2019.1666011
22. Kahn J. Learning at the Intersection of Self and Society: The Family Geobiography as a Context for Data Science Education // *Journal of the Learning Sciences*, 2020, 29 (1), С. 57–80.
23. Wise A. F. Educating Data Scientists and Data Literate Citizens for a New Generation of Data // *Journal of the Learning Sciences*, 2020, 29(1). С. 165–181.
24. Hardy L., Dixon C., Hsi S. From Data Collectors to Data Producers: Shifting Students' Relationship to Data // *Journal of the Learning Sciences*, 2020, 29(1). С. 104–126.
25. Cuoco A., Goldenberg E. P., Mark J. Habits of Mind: an organizing principle for mathematics curriculum // *Journal of Mathematical Behavior*, 1997, 15(4). С. 375–402.
26. Finzer, W. The data science education dilemma [Электронный ресурс]. // *Technology Innovations in Statistics Education*, 2013, 7(2). URL <https://escholarship.org/uc/item/7gv0q9dc> (дата обращения 10.04.2020).
27. De Veaux, R. и др. Curriculum guidelines for undergraduate programs in data science // *Annual Review of Statistics and Its Application*, 2017, 4(1). С. 15–30.
28. Hardin J. и др. Data science in statistics curricula: Preparing students to “think with data” // *The American Statistician*, 2015, vol. 69(4). С. 343–353.

Наука и инновации в информационном обществе

ЕДИНОЕ ЦИФРОВОЕ ПРОСТРАНСТВО НАУЧНЫХ ЗНАНИЙ В МИРОВОМ ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ

Статья рекомендована к публикации членом редакционного совета А.М. Елизаровым 30.09.2020.

Каленов Николай Евгеньевич

Д.т.н., профессор

Межведомственный суперкомпьютерный центр Российской академии наук – филиал Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», главный научный сотрудник

Москва, Россия

nekalenov@mail.ru

Соболевская Ирина Николаевна

К.ф.-м.н.

Межведомственный суперкомпьютерный центр Российской академии наук – филиал Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», старший научный сотрудник

Москва, Россия

ins@jicc.ru

Сотников Александр Николаевич

Д.ф.-м.н.

Межведомственный суперкомпьютерный центр Российской академии наук – филиал Федерального государственного учреждения «Федеральный научный центр Научно-исследовательский институт системных исследований Российской академии наук», заместитель директора

Москва, Россия

asotnikovov@mail.ru

Аннотация

Научное пространство является подпространством существующего информационного пространства, в котором содержится вся научная информация, представленная в различной форме (в виде печатных и электронных документов, аудио- и видеоматериалов, изображений, моделей пространственных объектов т.п.). В свою очередь, цифровое научное пространство также является подмножеством всего информационного пространства, в котором содержится научная информация в цифровой форме. Цифровое пространство знаний (ЦПЗ) – подмножество цифрового пространства, содержащее достоверную информацию. Единое цифровое пространство научных знаний (ЕЦПНЗ) – подмножество ЦПЗ, содержащее проверенную научную информацию, представленную в соответствии с единой онтологией. Под онтологией понимается свод правил построения и предоставления пользователям элементов ЕЦПНЗ. Поэтому, для построения ЕЦПНЗ, необходимо определить его элементы и онтологию их представления. В статье рассматривается структура ЕЦПНЗ как совокупности подпространств знаний по различным научным направлениям, виды и типы данных, составляющих контент ЕЦПНЗ, операции над ними, принципы управления пространством.

Ключевые слова

цифровое научное пространство, электронная библиотека, единое цифровое пространство, предметная онтология, научная сеть

© Каленов Н.Е., Соболевская И.Н., Сотников А.Н., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial - ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

Введение

Единое цифровое пространство научных знаний (ЕЦПНЗ) является одним из важнейших объектов современного информационного общества. Пространство в его математическом (формализованном) понимании есть множество некоторых объектов с определенными правилами работы с ними и набор аксиом, которым эти правила должны подчиняться, т.е. это множество с введенной на нем структурой. Определяя множество и структуру по-разному, можно получать разные пространства. Мировое информационное пространство содержит все сведения, накопленные человечеством в процессе его эволюции, представленные на материальных носителях. Оно включает различного рода документы, представленные в печатной, рукописной или электронной формах (публикации, архивные материалы, научно-техническую документацию и т.п.), фотографии, кино- видео- и аудиоматериалы, мультимедиа и 3D-модели объектов реального мира.

Цифровое пространство знаний (ЦПЗ), являющееся частью мирового информационного пространства, представляет собой компьютерную среду, содержащую формализованные характеристики разнородных знаний (это и законы механики, и химические формулы, и расписание движения транспорта, и рецепты приготовления различных блюд, и пр.). Цифровое пространство научных знаний (ЦПНЗ) – часть ЦПЗ, содержащая достоверную фундаментальную, научно-образовательную и научно-популярную информацию по различным областям науки, представленную в разнообразных формах. Единое цифровое пространство научных знаний (ЕЦПНЗ) – компьютерная среда, содержащая информацию, отраженную в ЦПНЗ, организованную и предоставляемую пользователям по единым правилам для всех областей науки (рис. 1).



Рисунок 1 – Информационные пространства

Иными словами, ЕЦПНЗ состоит из совокупности подпространств, относящихся к отдельным направлениям науки, связанных между собой на основе онтологии, общей для всего пространства [1, 2]. Общая онтология включает ряд предметных онтологий, описывающих с помощью тезаурусов и классификационных систем отдельные научные направления (рис. 2).

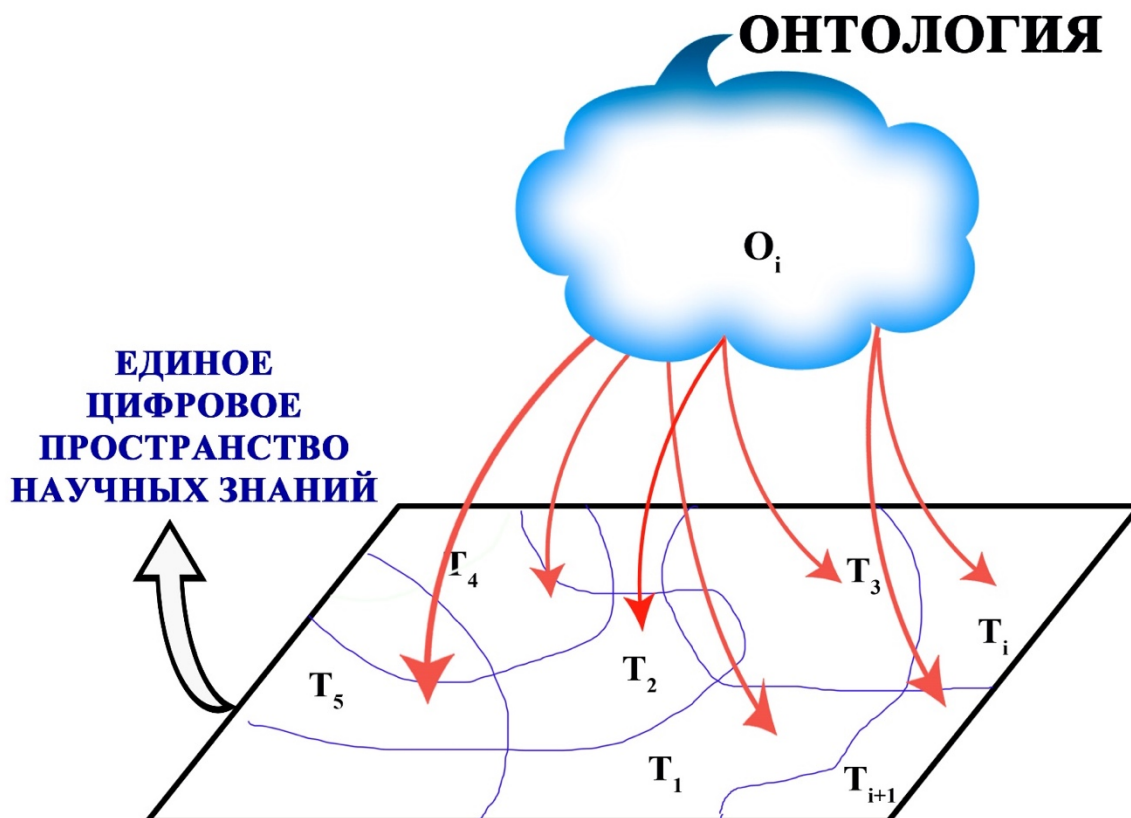


Рисунок 2. - Единое цифровое пространство научных знаний с общей онтологией.

Здесь

$T_1, \dots, T_i, T_{i+1}, \dots$ - тематические подмножества ЕЦПНЗ, представляющие собой ресурсы по различным областям науки (физике, математике, биологии и т.п.). Эти подмножества могут быть независимыми, а могут (в большинстве случаев) пересекаться.

O_i - множество предметных онтологий каждого тематического подпространства T_i

Каждое подпространство ЕЦПНЗ включает аксиомы и фундаментальные результаты, лежащие в основе данного научного направления, а также динамическую часть, содержащую информацию о новейших достижениях в данной научной области.

Для каждой отдельной области науки определены конкретные научные знания, но почти в любой области можно выделить два класса знаний - теоретические и экспериментальные [3].

Теоретические знания - совокупность аксиом, методов исследования и базовых результатов, полученных на основе применения этих методов. Аксиомы и методы исследований являются формальными и признаны научным сообществом со всеми устоявшимися элементами и понятиями [1]. При этом базовые результаты теоретических исследований - понятие, которое определяется экспертными оценками научного сообщества.

Экспериментальные знания - совокупность знаний и умений, приобретённых человеком в процессе взаимодействия с внешним по отношению к нему миром [1]. Эти знания основаны на результатах применения теоретических знаний, подтверждённых практикой. Очевидно, что такое деление относительно, поскольку множество теоретических знаний возникло в результате экспериментов (законы физики, механики и других естественных наук).

Источником наполнения ЕЦПНЗ являются ресурсы цифрового научного пространства, которые должны быть проанализированы на предмет достоверности, важности и актуальности.

В последнее время цифровое научное пространство активно пополняется ресурсами различного вида - формируются репозитории открытого доступа (<https://arxiv.org>) <https://ru.wikipedia.org/wiki/ArXiv.org> развиваются научные и образовательные социальные сети. Научные социальные сети предоставляют многочисленные сервисы по обмену разнообразной

информацией, размещению результатов исследований, их рецензированию и комментированию, поиску вакансий и т. п. [3].

Примерами научных социальных сетей могут служить такие сервисы, как:

Academia (www.academia.edu);

ResearchGate (www.researchgate.net);

Social Science Research Network (www.ssrn.com);

Mendeley (www.mendeley.com);

Citeulike (www.citeulike.org);

Myexperimen (www.myexperiment.org);

Соционет (<https://socionet.ru>);

Учёные России (www.russian-scientists.ru);

Научная сеть (www.scipeople.ru);

Социальная научная сеть (<https://www.science-community.org/>).

В настоящее время активно развиваются научные электронные библиотеки, предоставляющие широкому кругу пользователей такие сервисы, как:

Поиск метаданных документов по различным научным дисциплинам и источникам;

Полнотекстовый поиск в цифровых документах;

Предоставление доступа к найденным информационным ресурсам;

Обеспечение сохранности коллекций цифровых документов;

Примером такой научной электронной библиотеки, предоставляющей пользователю перечисленные выше сервисы, может служить Национальная цифровая библиотека по науке США (The National Science Digital Library – NSDL, <https://nsdl.oercommons.org>) [4, 5]. Этот ресурс основан Национальным научным фондом (NSF) в 2000 г. NSDL агрегирует ресурсы из цифровых библиотек и проектов, поддерживаемых NSF, а также формирует собственные обзоры вебсайтов по своим направлениям.

Среди достаточно развитых российских научных электронных библиотек необходимо выделить электронную библиотеку «Научное наследие России» (ЭБ ННР) [2, 3, 6]. Эта библиотека представляет собой объектно-ориентированную информационную систему, интегрирующую разнородные научные ресурсы институтов памяти (библиотек, архивов, музеев). Пользовательский блок ЭБ ННР (<http://e-heritage.ru>) предоставляет возможность многоаспектного поиска таких объектов как «персона», «публикация», «музейный предмет», «архивный документ», и навигации по найденным ресурсам.

Одним из недостатков электронных библиотек является поисковый сервис, не обеспечивающий достаточную точность и полноту поиска. В частности, в некоторых из них не предусмотрено использование при формировании запросов булевой логики.

Кроме электронных библиотек в интернете представлены различные коммерческие и некоммерческие научные сервисно-поисковые системы [7, 8]. Можно выделить следующие наиболее распространенные сервисно-поисковые системы:

Google Scholar (<https://scholar.google.ru/>) – поисковая система научной литературы;

Scholar.ru (<http://www.scholar.ru>) – поисковая система научных публикаций. Этот проект создан для упрощения поиска научных документов на русском языке. Основная цель этой поисковой системы – сбор информации о свободно скачиваемых научных публикациях;

Science Research Portal (<http://www.science-research.com/scienceresearch>) – научная поисковая система, осуществляющая полнотекстовый поиск в журналах многих крупных научных издательств, таких как Elsevier, Highwire, IEEE, Nature и др.;

CiteSeer Publications Research Index (<http://citeseer.ist.psu.edu/>) – научная поисковая система, индексирующая статьи в PostScript и PDF формате с научных веб-сайтов;

Ingenta (<http://www.ingentaconnect.com/>) – библиографическая база данных универсального типа, включающая описания статей из англоязычных журналов и сборников. Объем базы составляет более 13 млн записей [9].

Наиболее популярными источниками справочной информации являются, в частности, такие ресурсы, как Wikipedia и Wikidata. Эти ресурсы не верифицируются научным сообществом, поэтому не могут быть источником достоверных научных знаний.

Описанные выше примеры демонстрируют, что несмотря на то, что каждый из ресурсов мирового информационного пространства играет определённую роль, ни один из них не может претендовать на роль основы единого цифрового пространства научных знаний. Это связано, в том числе, с тем, что:

В большинстве случаев, имеющиеся ресурсы ориентированы на предоставление сведений о публикациях, которые содержат запрашиваемую информацию. В ряде случаев информационный сервис предоставляет полные тексты таких публикаций. Исключение составляют узкотематические фактографические системы, которые ориентированы на выдачу конкретных научных знаний. Они могут являться элементами соответствующих тематических подпространств, но не охватывают все виды объектов и не обеспечивают связи между подпространствами.

Контент политематических ресурсов (Wikipedia, Wikidata, Europeana) является случайным, и часто содержит неточную, непроверенную или неверную информацию [10].

Во всех системах отсутствует разделение контента по уровням на фундаментальные знания, научно-популярные и учебные материалы.

Формирование ЕЦПНЗ предполагает разработку специальных подходов и алгоритмов, которые базируются на новых принципах.

Структура ЕЦПНЗ

Пространство научных знаний должно включать две составляющие – статическую и динамическую. Статическая составляющая – это основополагающие, проверенные временем и практикой теоретические и экспериментальные данные. Динамическая составляющая – часть ЕЦПНЗ, охватывающая появляющиеся новые знания.

Эти составляющие можно рассматривать как две части пространства знаний, одна из которых (базис) отражает зафиксированные научные знания, другая (надстройка) – новую научную информацию. При этом, после прохождения экспертного фильтра, вторая часть переходит в первую.

Связи между базисом и надстройкой ЕЦПНЗ могут быть реализованы, в том числе, на уровне междисциплинарной общенаучной онтологии [1]. При этом, базис и надстройка представляют собой совокупность подпространств (фасетов) по различным научным направлениям.

Базис включает три взаимосвязанных уровня:

Образовательный, рассчитанный, в частности, на учащихся школ. Этот уровень статический, формируется на основе информации, полученной из академической литературы, например, учебников, научно-популярной литературы. На этом же уровне могут быть представлены верифицированные мультимедийные ресурсы, например, оцифрованные музейные объекты.

Научно-популярный, рассчитанный на более «продвинутых» пользователей, но не специалистов в данной области. Этот уровень квазистатический. Он может изменяться при появлении новых существенных для данной области знания открытий или результатов. Формирование этого уровня строится на основе энциклопедической и справочной информации, а также монографий и учебников для высшей школы. На этом уровне наряду с извлечёнными метаданными, пользователь может получить ссылки на источники и их цифровые копии в соответствии с законом об охране авторских прав.

Фундаментальный, предназначенный для специалистов в данной области науки. Этот уровень включает более глубокую по сравнению с предыдущими двумя уровнями информацию по данному научному направлению. Которая, также является статической. Контент на этом уровне может быть дополнен и подтвержден экспертными оценками новой информацией.

На рисунке 3 показано распределение новой информации по трем базисным уровням.



Рисунок 3 – Распределение новой информации по трем базисным уровням

Надстройка, ориентированная на специалистов в данной области науки, является источником текущего контента, необходимого для проведения научных исследований. Она содержит новые результаты в той или иной научной области. Контент надстройки формируется за счёт новой информации, прошедшей экспертизу, например, информации из статей, опубликованных в верифицированных журналах, из сборников научных статей с подтверждённым научным уровнем, патентов, материалов, ведущих мировых научных конференций и пр. Наполнение надстройки, динамически меняется. Это происходит по причине исключения информации, ценность которой не подтвердилась временем, или благодаря новой, проверенной информации.

Контент ЕЦПНЗ

Информационные ресурсы являются источниками научных знаний. Они содержат описания постулатов, теорий, экспериментов, полученных на их основе результатов в той или иной области науки, представленные на материальном носителе. Материальные носители эволюционировали от глиняных табличек и папирусов до печатных, фото-, кино-, видеоматериалов и цифровых музейных моделей [11-15]. Современные цифровые носители обладают качественно иными возможностями с точки зрения компактного хранения, извлечения, поиска и предоставления знаний пользователям.

Базис ЕЦПНЗ формируется из имеющихся библиотечных, архивных, музейных, энциклопедических, фактографических, словарных и других ресурсов. Этот контент создается научно-образовательным сообществом на основе существующих систем без нарушения их функционирования. Как правило, информация, содержащаяся в этих ресурсах, является достоверной и проверенной временем. Тем не менее, для принятия решения о том, что будет загружено в ЕЦПНЗ, необходима экспертиза. Экспертами должны выступать квалифицированные представители научного сообщества в области, соответствующей данному подпространству.

Динамическая часть (надстройка) ЕЦПНЗ формируется информационно-библиотечными специалистами совместно с учёными на основе существующих мировых баз данных научной информации и коллекций полнотекстовых цифровых документов.

Базис и надстройка ЕЦПНЗ состоят из ядра и оболочки.

Элементы ядра представляют собой информационные объекты, которые содержат метаданные, полученные из перечисленных выше ресурсов, определяющих основные постулаты, методы и результаты исследований в конкретной области науки для базиса и надстройки.

Каждый информационный объект должен:

Содержать элементы типа «Что (кто), где, когда»;

Указывать на связи с другими объектами как данного подпространства, так и, в общем случае, других подпространств;

Иметь ссылку на источник информации.

Оцифрованные публикации, архивные материалы, образы музейных экспонатов, мультимедийные материалы, тематические базы данных, поддерживаемые в научных организациях, образуют *оболочку ЕЦПНЗ*.

Предметная онтология ЕЦПНЗ

Предметные онтологии различных научных областей широко применяются в задачах построения поисковых систем, систем представления знаний, а также при решении различного рода семантических задач [16-18]. Процедура введения онтологических представлений в накопленный массив данных представляет собой определение предметной области через описание множества используемых в ней понятий и связей между ними [19]. Одной из задач формирования ЕЦПНЗ является разработка платформы, которая позволит автоматизировать процесс построения связей между объектами предметной области. Эта платформа может выступать в роли независимой информационной среды, а может быть внедрена в качестве одного из элементов в более масштабную информационную среду, создаваемую, в том числе, для образовательного процесса и научных исследований. Одной из важнейших задач при построении ЕЦПНЗ является формирование обобщенной предметной онтологии, включающей единые тезаурусы общенаучных терминов, а также словари номенклатурных наименований, географических названий, имён персоналий, событий, обозначения временных интервалов, тематики и другие категории научной лексики. Обобщенная онтология должна строиться на базе предметных онтологий, которые создаются внутри каждого подпространства, отражающего конкретное научное направление. Эти онтологии включают ключевые слова и корреляции между ними, а также средства, обеспечивающие взаимодействие различных систем классификации научных знаний в той или иной научной области.

Унифицированное представление связей между объектами, единые принципы описания информационных объектов, единые терминологические словари и тезаурусы, единые принципы оцифровки источников информации – поддержка всего этого обеспечивает единство пространства научных знаний.

Операции над объектами ЕЦПНЗ

Для ЕЦПНЗ определены следующие операции над информационными объектами:

Поиск объектов по любым элементам метаданных и их комбинациям в каждом из подпространств;

Поиск объектов по любым элементам метаданных и их комбинациям в цифровом пространстве в целом;

Выдача списков найденных объектов;

Сортировка найденных объектов по заданному элементу;

Навигация от найденных объектов к связанным с ними объектам, находящимися, как внутри подпространства, так и в пространстве в целом с использованием возможностей общей онтологии ЕЦПНЗ;

Формирование различных коллекций объектов, отвечающих заданному параметру.

Для числовых значений метаданных определены операции:

«равно»;

«не равно»;

«меньше»;

«больше».

Для текстовых объектов определены операции:

«равно» – это значение предполагает точное совпадение элемента метаданных с заданным фрагментом текста с возможностью выбора значения из перечислимого множества);

«начинается с...»;

«содержит»;

«не содержит»;

правое и левое усечения от начала слова.

Для данных типа «дата» определены операции:

«равно»;

«до заданной даты»;

«после заданной даты»;

«в заданном интервале».

Поисковый запрос может содержать множество элементов, связанных операторами:

«и»;

«или»;

«и не».

Развитие ЕЦПНЗ предполагает расширение наполнения и увеличение номенклатуры входящих в него объектов, что, в свою очередь, повлечет появление новых запросов и потребует введения новых операций на множестве объектов ЕЦПНЗ.

Управление ЕЦПНЗ

Единое пространство знаний формируется на основе распределённой (децентрализованной) подготовки метаданных. Эти метаданные должны создаваться по единым правилам с централизованным хранением при единой централизованной системе контроля качества. Материалы, относящиеся к частям ЕЦПНЗ (оцифрованные публикации, архивные документы, музейные предметы и т. п.), могут храниться у владельцев этих данных или в центральном блоке пространства.

С функциональной точки зрения управление ЕЦПНЗ включает три уровня [20]:

Верхний уровень управления обеспечивает центральный блок, связанный через обобщенную онтологию с блоками управления ядрами тематических подпространств ЕЦПНЗ, который, в свою очередь связан с блоками управления данными оболочек подпространств (рис. 4).

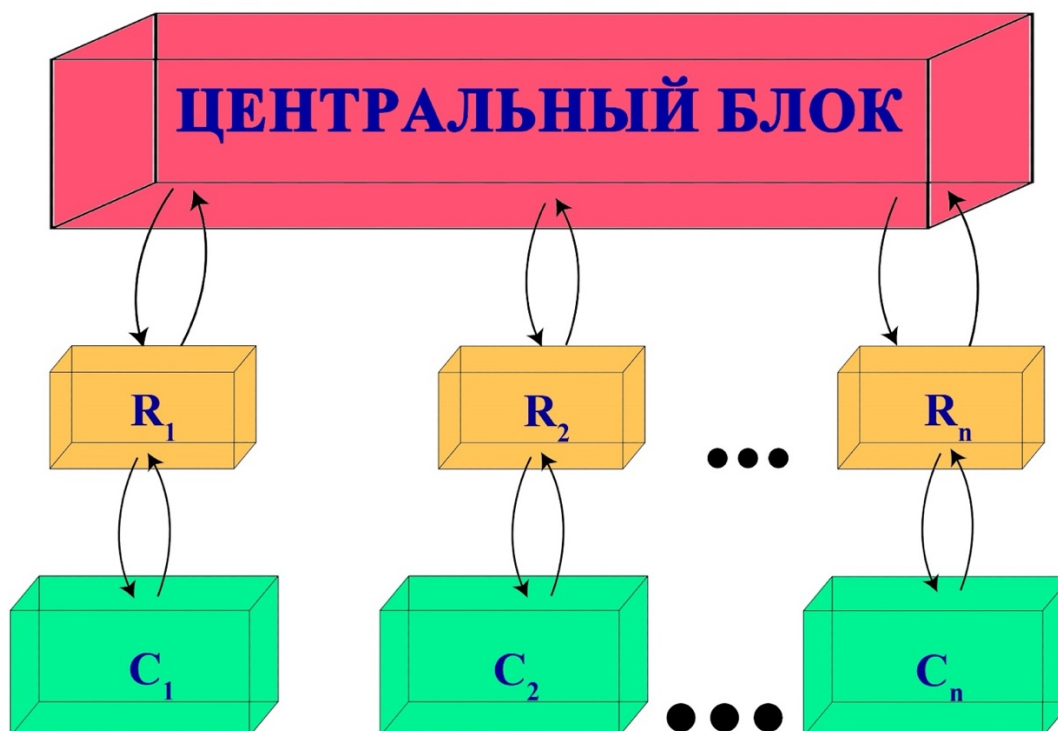


Рисунок 4. Структура управления ЕЦПНЗ

Здесь

$R_1, R_2 \dots R_n$ – блоки управления ядрами тематических подпространств ЕЦПНЗ;

$C_1, C_2 \dots C_n$ – блоки управления оболочками тематических подпространств ЕЦПНЗ.

Блоки управления ядрами обеспечивают загрузку и поддержку содержимого ядра каждого подпространства. Блоки управления оболочками – обеспечивают загрузку цифровых объектов и их связи с ядрами.

Заключение

Единое цифровое пространство научных знаний должно создаваться усилиями научных организаций, библиотек, музеев и архивов. При этом необходим мониторинг существующих ресурсов; определение принципов отбора наиболее значимых объектов, разработка соответствующей программной оболочки и технологии наполнения контента. Необходимо разработать общесистемную онтологию и начать поэтапную реализацию отдельных тематических подпространств. Работы по перечисленным выше направлениям ведутся в настоящее время в МСЦ РАН в кооперации с другими академическими институтами и научными библиотеками в рамках государственного задания, а также при поддержке РФФИ.

Работа выполнена в МСЦ РАН – филиале ФГУ ФНЦ НИИСИ РАН рамках государственного задания 065-2019-0014.

Литература

1. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н., 2019. О едином цифровом пространстве научных знаний. Вестник Российской Академии Наук. Т. 89(7), сс. 728-735.
2. Антопольский А.Б., 2019. О разработке Единого российского электронного пространства знаний. Информации и инновации. Т. 14(1), сс. 7-17.
3. Антопольский А.Б., Ефременко Д.В., 2018. К вопросу о едином электронном пространстве знаний. Вестник Российской Академии Наук. Т. 88(1), сс. 89-95.

4. Andrews-Todd, J., Forsyth Carol M., 2018. Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. In *Computers in human behaviour*. Vol. 104, UNSP 105759.
5. Arshad H., Jantan A., Hoon G.K., Abiodun I.O., 2020. Formal knowledge model for online social network forensics. In *Computers & security*. Vol. 89, UNSP 101675.
6. Погорелко К.П., 2015. Анализ спроса на электронную библиотеку «научное наследие России». Информационное обеспечение науки: новые технологии. Сборник научных трудов, сс. 191-199.
7. Zhmailo S.V., Ulyanin O.V.. Sci-tech libraries within the knowledge management system: from information specialist's viewpoint. 2020. In *Nauchnye i tekhnicheskie biblioteki-scientific and technical libraries*. vol. 2, pages 9–23.
8. Sanchez A.S., 2017. Knowledge Organization Systems: Definition and Historical Development. In *E-ciencias de la information*. vol. 7(2).
9. Corrales D.C . Ledezma, A., Corrales, J.C., 2020. A case-based reasoning system for recommendation of data cleaning algorithms in classification and regression tasks. In *Applied soft computing*. Vol. 90, UNSP 106180.
10. Deena G., Raja K., 2019. Designing an Automated Intelligent e-Learning System to Enhance the Knowledge using Machine Learning Techniques. I n *International journal of advanced computer science and applications*. Vol. 10(12), pages 112-119.
11. Dresel R., Henkel M., Scheibe K., Zimmer F., Stock WG., 2020. A Nationwide Library System and Its Place in Knowledge Society and Smart Nation: The Case of Singapore. In *LIBRI*. vol. 70(1), pages 81-94.
12. Frandsen T.F., Tibyampansha D., Ibrahim G.R., von Isenburg M., 2017. Library training to promote electronic resource usage: A case study in information literacy assessment. In *Information and Learning Science*. vol. 118(11-12), pages 618-628.
13. Gabrys J., 2020. Smart forests and data practices: From the Internet of Trees to planetary governance. In *Big data & society*. Vol. 7(1), UNSP 2053951720904871.
14. Gushul Y.V., Teslya E.V., 2020. Information and analytical support: current goals and pathways. In *Nauchnye i tekhnicheskie biblioteki-scientific and technical libraries*, pages 24-44.
15. Maggio A., Kuffer J., Lazzari M., 2017. Advances and trends in bibliographic research: Examples of new technological applications for the cataloging of the georeferenced library heritage. In *Journal of librarianship and information science*. vol. 49(3), pages 299-312.
16. Yeh J.H., Huang X.M., 2018. BKOntoVR: A Virtual Reality Exhibition System for Biographic Ontology-Based Semantic Structure. In *Proceedings of 2018 2nd international conference on software and e-business*, pages 69-73.
17. Lou W., Pi RF., Wang H., Ju Y., 2019. Low-cost similarity calculation on ontology fusion in knowledge bases. In *Journal of information science*. SCITEPRESS.
18. Bonacchi C., Krzyzanska M., 2019. Digital heritage research re-theorised: ontologies and epistemologies in a world of big data. In *International journal of heritage studies*. Vol. 25(12), 1235-1247.
19. Shahzad F., Alwosaibi F.M. 2017. Development of an e-Library Web application. IMSCI. In *11th International Multi-Conference on Society; Orlando; the United States, Cybernetics and Informatics*. Pages 153-158.
20. Kalenov N.E., Sobolevskaya I.N., Sotnikov A.N., 2019. Hierarchical representation of information objects in a digital library environment. In *17th Russian Conference, RCAI 2019, Ulyanovsk, Russia*. SCITEPRESS.

COMMON DIGITAL SPACE OF SCIENTIFIC KNOWLEDGE IN THE WORLD INFORMATION SPACE

Kalenov, Nikolai Evgenievich

Doctor of technical sciences, professor

*Joint Supercomputer Center of the Russian Academy of Sciences – Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences”, chief researcher
Moscow, Russia
nekalenov@mail.ru*

Sobolevskaya, Irina Nikolaevna

Candidate of physical and mathematical sciences

*Joint Supercomputer Center of the Russian Academy of Sciences – Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences”, senior researcher
Moscow, Russia
ins@jsc.ru*

Sotnikov, Alexander Nikolaevich

Doctor of physical and mathematical sciences

*Joint Supercomputer Center of the Russian Academy of Sciences – Branch of Federal State Institution “Scientific Research Institute for System Analysis of the Russian Academy of Sciences”, deputy director
Moscow, Russia
asotnikovov@mail.ru*

Abstract

The scientific space is a subspace of the existing information space. The scientific space contains all the scientific information presented in various forms (printed sources, digital documents, audio and video materials, images, 3D-models of various objects, etc.). In turn, the digital scientific space is also a subset of the entire information space, which contains scientific information in digital form. Digital Knowledge Space (DKS) is a subset of the digital space that contains reliable information. The Common digital space of scientific knowledge (CDSSK) is a subset of the DKS containing verified scientific information presented in accordance with a single ontology. Ontology is understood as a set of rules for constructing and providing users with the elements of the CDSSK. Therefore, in order to build the CDSSK, it is necessary to determine its elements and the ontology of their presentation. The article discusses the structure of the CDSSK as a set of knowledge subspaces in various scientific areas, the types and types of data that make up the content of the CDSSK, operations with them, the principles of space management.

Keywords

digital knowledge space, information space, ontology, digital library, subject ontologies

References

1. Antopol'skij A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N., 2019. O edinom cifrovom prostranstve nauchnyh znaniy. Vestnik Rossijskoj Akademii Nauk. T. 89(7), ss. 728-735.
2. Antopol'skij A.B., 2019. O razrabotke Edinogo rossijskogo elektronnoho prostranstva znaniy. Informacii i innovacii. T. 14(1), ss. 7-17
3. Antopol'skij A.B., Efremenko D.V., 2018. K voprosu o edinom elektronnom prostranstve znaniy. Vestnik Rossijskoj Akademii Nauk. T. 88(1), ss. 89-95.
4. Andrews-Todd, J., Forsyth Carol M., 2018. Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. In Computers in human behaviour. Vol. 104, UNSP 105759.
5. Arshad H., Jantan A., Hoon G.K., Abiodun I.O., 2020. Formal knowledge model for online social network forensics. In Computers & security. Vol. 89, UNSP 101675.
6. Pogorelko K.P., 2015. Analiz sprosa na elektronnyyu biblioteku «nauchnoe nasledie Rossii». Informacionnoe obespechenie nauki: novye tekhnologii. Sbornik nauchnyh trudov, ss. 191-199.

7. Zhmailo S.V., Ulyanin O.V.. Sci-tech libraries within the knowledge management system: from information specialist's viewpoint. 2020. In *Nauchnye i tekhnicheskie biblioteki-scientific and technical libraries*. vol. 2, pages 9–23.
8. Sanchez A.S., 2017. Knowledge Organization Systems: Definition and Historical Development. In *E-ciencias de la information*. vol. 7(2).
9. Corrales D.C . Ledezma, A., Corrales, J.C., 2020. A case-based reasoning system for recommendation of data cleaning algorithms in classification and regression tasks. In *Applied soft computing*. Vol. 90, UNSP 106180.
10. Deena G., Raja K., 2019. Designing an Automated Intelligent e-Learning System to Enhance the Knowledge using Machine Learning Techniques. In *International journal of advanced computer science and applications*. Vol. 10(12), pages 112-119.
11. Dresel R., Henkel M., Scheibe K., Zimmer F., Stock WG., 2020. A Nationwide Library System and Its Place in Knowledge Society and Smart Nation: The Case of Singapore. In *LIBRI*. vol. 70(1), pages 81-94.
12. Frandsen T.F., Tibyampansha D., Ibrahim G.R., von Isenburg M., 2017. Library training to promote electronic resource usage: A case study in information literacy assessment. In *Information and Learning Science*. vol. 118(11-12), pages 618-628.
13. Gabrys J., 2020. Smart forests and data practices: From the Internet of Trees to planetary governance. In *Big data & society*. Vol. 7(1), UNSP 2053951720904871.
14. Gushul Y.V., Teslya E.V., 2020. Information and analytical support: current goals and pathways. In *Nauchnye i tekhnicheskie biblioteki-scientific and technical libraries*, pages 24-44.
15. Maggio A., Kuffer J., Lazzari M., 2017. Advances and trends in bibliographic research: Examples of new technological applications for the cataloging of the georeferenced library heritage. In *Journal of librarianship and information science*. vol. 49(3), pages 299-312.
16. Yeh J.H., Huang X.M., 2018. BKOntoVR: A Virtual Reality Exhibition System for Biographic Ontology-Based Semantic Structure. In *Proceedings of 2018 2nd international conference on software and e-business*, pages 69-73.
17. Lou W., Pi RF., Wang H., Ju Y., 2019. Low-cost similarity calculation on ontology fusion in knowledge bases. In *Journal of information science*. SCITEPRESS.
18. Bonacchi C., Krzyzanska M., 2019. Digital heritage research re-theorised: ontologies and epistemologies in a world of big data. In *International journal of heritage studies*. Vol. 25(12), 1235-1247.
19. Shahzad F., Alwosaibi F.M. 2017. Development of an e-Library Web application. IMSCI. In *11th International Multi-Conference on Society; Orlando; the United States, Cybernetics and Informatics*. Pages 153-158.
20. Kalenov N.E., Sobolevskaya I.N., Sotnikov A.N., 2019. Hierarchical representation of information objects in a digital library environment. In *17th Russian Conference, RCAI 2019, Ulyanovsk, Russia*. SCITEPRESS.

Информационное общество и право

СОДЕРЖАНИЕ ПРАВА НА ИНФОРМАЦИЮ В КОНТЕКСТЕ РЕЖИМА ИНФОРМАЦИИ ОТКРЫТОГО И ОГРАНИЧЕННОГО ДОСТУПА

Статья рекомендована к публикации членом редакционного совета А.А. Стрельцовым 25.12.2020.

Лясковский Иван Иванович

Соискатель степени PhD юридических наук

Государственный таможенный комитет Республики Беларусь, Витебская таможня, главный инспектор юридического отдела

Витебск, Республика Беларусь

ivanl@inbox.ru

Аннотация

Право на информацию представляется как упорядоченная система правомочностей, а также ограничений и запретов, применяемых в связи с обеспечением баланса интересов человека, общества и государства как участников информационных отношений. В рамках правового режима открытой информации рассматриваются активная и пассивная формы реализации права на информацию. В рамках режима информации ограниченного доступа – правовые ценности, обеспечение которых обосновывает возможность ограничения права на информацию.

Ключевые слова

правовые режимы информации; открытая информация; информация ограниченного доступа; законодательство Республики Беларусь; информационное право; право на информацию

Введение

Право на информацию является неотъемлемым элементом системы основополагающих прав и свобод человека и гражданина. Реализация этого права, в наше время, является необходимым условием участия субъектов права в общественных отношениях, а также – обеспечения гармоничного свободного развития и жизнедеятельности каждого человека. Принимая во внимание мнение Конституционного Суда Республики Беларусь о «становлении информационной сферы как системообразующего фактора жизни людей», можем констатировать, что юридическое содержание информационных прав человека выходит за пределы «формальной» совокупности правомочий. Оно приобретает свое значение как приоритетный жизненно-важный интерес, обеспечение которого имеет определяющее влияние на формирование административно-правовых режимов.

За последние десятилетия термин «информационные права человека» получил широкое распространение в правовой науке. Но вместе с тем, разные исследователи, характеризуя их содержание, приводят неодинаковые перечни прав. Причиной такого состояния является то, что права человека в сфере информации могут предполагать:

- 1) возможность получения, использования и распространения информации;
- 2) охрану частной жизни, защиту чести и достоинства, которые – в связи с применением ограничительных мер – могут показаться противоположными по содержанию праву на доступ к информации;
- 3) право на доступ к публичной информации, право на защиту персональных данных, право на доступ к интернету.

© Лясковский И.И., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция – Некоммерческое использование – На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

Соотношение права на информацию и информационных прав

Анализ научных источников не дает однозначного ответа на вопрос о соотношении права на информацию и информационных прав, но позволяет выделить два имеющихся научных подхода. Согласно первому подходу, «информационные права» объединяют:

- 1) право на доступ к информации (в его нормативной формулировке);
- 2) право на свободу мысли и слова, право на свободное выражение своих взглядов и убеждений и т.д.

В Конституции Республики Беларусь использованы каждый из этих подходов. Так, в нормативной конструкции ст. 33 Основного Закона говорится о гарантиях «свободы мнений, убеждений и их свободного выражения», а в ст. 34 о праве граждан Беларуси «получать, хранить и распространять полную, достоверную и своевременную информацию о деятельности государственных органов, общественных объединений, о политической, экономической, культурной и международной жизни, состоянии окружающей среды».

Учитывая то, что в основных структурных элементах права на информацию в ч. 1 ст. 34 Конституции Республики Беларусь закреплены такие правовые возможности как: право получать, право хранить, право распространять информацию, а в ч. 2 ст. 16 Закона Республики Беларусь от 10 ноября 2008 г. № 455-З «Об информации, информатизации и защите информации» приведен широкий перечень информации, доступ к которой не может быть ограничен – можем заключить, что право на информацию является комплексным.

Для второго подхода характерно определенное смысловое отождествление права на информацию и информационных прав, основанное на современном восприятии права на информацию как определенным образом обособленного личного неимущественного права физического лица, а сквозь призму науки информационного права – как универсального конституционного права на информацию, что включает в себя комплекс возможностей, которые в совокупности и составляют так называемые информационные права субъектов.

Так, Д.В. Огородов к субъективным информационным правам относит абсолютные права на информацию (например, право на защиту информации, составляющей личную тайну) [4, с. 15-16].

А. Малько считает, что право на информацию охватывает такие конкретные правомочия:

- 1) право знать о создании и функционировании всех конкретных информационных систем, которые в какой-либо степени затрагивают сферу личной жизни гражданина или информацию о ней, а также информацию об иных сферах жизнедеятельности гражданина;
- 2) право давать согласие на сбор личностной информации для социально-экономических, культурных и иных социальных целей;
- 3) право проверять достоверность такой информации и оспаривать недостоверную информацию как в административном, так и в судебном порядке;
- 4) право доступа к такой информации с целью ее проверки, получения необходимых справок;
- 5) право знать об использовании этой информации в соответствующих целях и соответствующими пользователями систем;
- 6) право на гражданский код (обозначение гражданина в соответствующей информационной системе);
- 7) право на достоверную информацию о состоянии окружающей природной среды;
- 8) право на достоверную финансовую информацию, и другие [3, с. 60].

Попытка сравнения приведенных подходов в действующем информационном законодательстве, позволяет сделать вывод о том, что эти подходы являются взаимодополняющими. С одной стороны, видим, что право на информацию находится в системе основных прав и свобод человека и гражданина, поэтому вполне точно можно говорить о его взаимосвязи и взаимовлиянии с другими правами человека. Например, право на информацию имеет тесную и безоговорочную связь с такими конституционными правами граждан как: участие в управлении предприятиями, организациями и учреждениями (ч. 8 ст. 13 Конституции Республики Беларусь); участие в решении государственных дел (ч. 1 ст. 37); получение образования (ч. 1 ст. 49 Конституции), и другими.

С другой стороны, механизм реализации информационных прав требует установление правовых режимов для разных видов информации. Режимное регулирование позволяет учитывать

особенности осуществления информационных прав человека сообразно специфике конкретной сферы правоотношений, а также обеспечивает полноту реализации законных интересов граждан, их сбалансированность с интересами общества и государства.

Реализацию права на информацию в контексте правовых режимов информации открытого или ограниченного доступа следует рассматривать в связи с активной и пассивной формой реализации права на информации.

Право на информацию можем рассматривать в его активной форме, реализуемой путем непосредственного участия граждан в информационных отношениях. Согласно абз. 17 ч. 1 ст. 1 Закона Республики Беларусь от 10 ноября 2008 г. № 455-З «Об информации, информатизации и защите информации», информационные отношения – это отношения, возникающие при поиске, получении, передаче, сборе, обработке, накоплении, хранении, распространении и (или) предоставлении информации, пользовании информацией, защите информации, а также при применении информационных технологий.

Реализация права на информацию в его пассивной форме корреспондирует с обязательствами некоторых субъектов частного и публичного права: обеспечить доступность определенных сведений (например, в ст. 7 Закона Республики Беларусь от 9 января 2002 г. № 90-З «О защите прав потребителей» приведены перечни данных, которые должен обнародовать производитель товара) или обеспечить осведомленность целевого пользователя информации по определенному вопросу (например, согласно п. 4 ч. 1 ст. 44 Избирательного кодекса Республики Беларусь от 11 февраля 2000 г. № 370-З участковые комиссии обязаны информировать избирателя о дне, времени и месте проведения выборов).

Активная и пассивная форма реализации права на информацию имеет определенные особенности в связи с гражданством лица (или отсутствием гражданства). Здесь можем рассматривать:

- 1) открытую информацию, к которой могут иметь беспрепятственный доступ любые лица;
- 2) открытую информацию, получение которой государство обязано обеспечить только для граждан государства (с целью выполнения ими своих обязанностей или реализации прав);
- 3) информацию с ограниченным доступом, к которой могут получить доступ только граждане государства (например, доступ к государственным секретам после выполнения лицом ряда условий, которые выдвигаются Законом Республики Беларусь от 19 июля 2010 г. № 170-З «О государственных секретах» для получения такого доступа).

Также свои особенности имеет порядок распространения информации в связи с возрастом ее потенциальных пользователей. Так, ст. 37-2 Закона Республики Беларусь от 19 ноября 1993 г. № 2570-ХП «О правах ребенка» установлено, что запрещается распространение информационной продукции (согласно приведенному в этой статье перечню видов продукции – И.Л.) «...без присвоения знака возрастной категории ... указывающего на возрастную категорию детей, среди которых допускается распространение информационной продукции...».

Через закрепление обязанности производителей присваивать возрастную категорию своей продукции происходит обеспечение *пассивного* права родителей на воспитание детей в безопасной информационной среде. Вместе с тем следует отметить, что реализация детьми определенной возрастной группы *активного* права на информацию – путем самостоятельного выбора того или иного информационного продукта – может осуществляться как с учетом, так и с игнорированием установленных для продукта возрастных ограничений.

В связи с этим Конституционный Суд Республики Беларусь говорит о том, что «... конституционно-правовой статус ребенка обязывает производителей и распространителей информационной продукции формировать стандарты благоприятной информационной среды, сводя к минимуму распространение материалов ... способных оказать негативное влияние на здоровье, физическое, нравственное и духовное развитие ребенка».

Таким образом, существует механизм возникновения обязательств субъектов частного и публичного права в связи с обеспечением права на информацию в рамках правового режима информации открытого доступа. Этот пример показывает, что обязательства могут возникать из активного права человека на доступ к информации – и заключаться в обнародовании и обеспечении доступности такой информации, а также из пассивного права человека на пребывание в благоприятной окружающей среде (ст. 46 Конституции Республики Беларусь) – и заключаться в обеспечении безопасности информационной среды.

Реализация права на информацию в контексте правового режима открытой информации показывает, что люди и государство взаимодействуют через выполнение гарантированных и охраняемых законом прав и обязанностей, закрепленных в Основном Законе и нормативных правовых актах в сфере информации. Определена их субъектность в информационных отношениях, предусмотрены основания и содержание юридической ответственности друг перед другом за нарушение прав. Таким образом, достигается разумный баланс между интересами в сфере открытой информации, присущих отдельным лицам, обществу и государству. Определяя права человека в сфере информации, государство ориентируется прежде всего на социально-типичные характеристики личности, создает систему взаимных прав и обязанностей, определяющих ее правосубъектность в этой сфере.

Рассматривая проблематику права на информацию, нельзя обойти вниманием вопрос его ограничения. Речь идет не об определении юридических границ естественного права, а об ограничении прав, закрепленных в соответствующих нормативных актах. Согласно ст. 23 Конституции Республики Беларусь ограничение права на информацию является правомерным только на основании законодательного акта, предусматривающего такую санкцию, и только в случаях, когда подобного ограничения требуют:

- 1) интересы национальной безопасности;
- 2) охрана общественного порядка;
- 3) защита нравственности;
- 4) защита здоровья населения;
- 5) защита прав и свобод других лиц.

Приведенные критерии выполняют роль универсальных правовых ценностей, которыми могут обосновываться ограничения права на информацию в рамках административного регулирования деятельности субъектов права. Например, в Решении Конституционного Суда Республики Беларусь от 10 июля 2018 г. № Р-1135/2018 «О соответствии Конституции Республики Беларусь Закона Республики Беларусь «О внесении изменений и дополнений в некоторые законы Республики Беларусь» изложена позиция, согласно которой регулирование деятельности СМИ в Республике Беларусь должно осуществляться как с учетом роли СМИ в гражданском обществе, так и с учетом необходимости соблюдения баланса конституционных прав и свобод личности и публичных интересов государства и общества.

Некоторые ограничения права на информацию в связи с деятельностью СМИ – могут быть правомерными, если основываются на необходимости обеспечения информационной безопасности как составной части национальной безопасности Республики Беларусь, повышения качества теле- и радиопрограмм и др. В качестве допустимых ограничений рассматривались такие нормы Закона Республики Беларусь от 17 июля 2018 г. № 128-З «О внесении изменений и дополнений в некоторые законы Республики Беларусь»: п. 3 ст. 2 о перечне лиц, которые не могут выступать в качестве учредителя средства массовой информации; п.9 ст. 2 об основаниях для отказа в государственной регистрации (перерегистрации) средства массовой информации; п. 22 ст. 2 о перечнях сведений, распространение которых в средствах массовой информации, на интернет-ресурсах ограничено; п. 28 ст. 2 об условиях ограничения доступа к интернет-ресурсу, сетевому изданию, и др.

В Решении Конституционного Суда Республики Беларусь от 26 декабря 2013 г. № Р-886/2013 «О соответствии Конституции Республики Беларусь Закона Республики Беларусь «О внесении изменений и дополнений в Закон Республики Беларусь «Об информации, информатизации и защите информации» дано заключение о том, что «ограничение распространения и (или) предоставления информации ограничивает и гарантированное Конституцией право на информацию. Однако указанное ограничение согласуется с положениями ч. 1 ст. 23 Конституции Республики Беларусь по форме закрепляющего его акта и направлено на защиту интересов национальной безопасности, общественного порядка, защиты нравственности, здоровья населения, прав и свобод других лиц».

Также Суд отметил, что законодательное регулирование, в т.ч. через ограничения, призвано усилить информационную безопасность граждан, в рамках реализации ст. 28 (о защите от незаконного вмешательства в личную жизнь) и ст. 34 (о защите чести, достоинства, личной и семейной жизни граждан и полного осуществления ими своих прав) Конституции Республики Беларусь. При этом отмечается, что ограничительные положения Закона Республики Беларусь от 10 ноября 2008 г. № 455-З «Об информации, информатизации и защите информации» согласуются

с нормами международно-правовых актов, что отвечает конституционному положению о признании Республикой Беларусь приоритета общепризнанных принципов международного права и обеспечении соответствия им законодательства.

Рассматривая границы допустимого вмешательства в тайну личной жизни, также обратимся к судебной практике. В Решении Конституционного Суда Республики Беларусь от 9 июля 2019 г. № Р-1187/2019 «О соответствии Конституции Республики Беларусь Закона Республики Беларусь «Об изменении Закона Республики Беларусь «Об аудиторской деятельности» высказано положение о том, что «... поскольку ограничение прав и свобод личности, в том числе в целях защиты прав и свобод других лиц, в силу ст. 23 Конституции Республики Беларусь допускается только в случаях, предусмотренных законом, то реализация конституционного права на информацию, затрагивающую частную жизнь других лиц, возможна только в порядке, установленном законом, и что только законодатель в силу предоставленной ему дискреции правомочен определить законные способы получения такой информации, установить в отношении той или иной информации специальный правовой режим, в том числе режим ограничения свободного доступа к ней».

Устанавливая такой баланс в отношении информации о частной жизни лиц, занимающихся аудиторской деятельностью, Суд признал обоснованным размещение определенных сведений о таких лицах в аудиторском реестре.

По мнению Суда, публикация таких сведений «... направлена на реализацию таких принципов аудиторской деятельности, как профессиональная компетентность и профессиональное поведение, для того чтобы аудиторские услуги оказывались аудиторами, обладающими необходимой профессиональной квалификацией, проявляющими непредвзятость и беспристрастность при рассмотрении профессиональных вопросов, формировании выводов и мнений, не совершающими дискредитирующих действий».

Вышеприведенные примеры позволяют согласиться с тезисом И.А. Ильина о том, что ограничения играют огромную роль в получении личностью подлинной свободы. Ограничивая права и свободы каждого индивида, закон обеспечивает ему беспрепятственное использование своих прав, то есть гарантирует свободу внутри этих границ. Свобода каждого человека распространяется только до того предела, от которого начинается свобода других людей. Стремясь установить эти границы, закон способствует тому, чтобы в совместной жизни людей воцарился порядок, основанный на свободе [2, с. 96].

Реализацию права на информацию в контексте информации ограниченного доступа можем рассматривать в связи с положительными и негативными обязанностями субъектов права, которые возникают в пределах того или иного правового режима информации.

Негативные права человека предусматривают обязательства государства и его агентов (представителей) воздерживаться от любых действий, направленных на их нарушение или незаконное ограничение [1, с. 335-336]. Это может выражаться в гарантиях невмешательства в личную жизнь человека, в охране тайны частной жизни и сведений о ней, в охране тайны телефонных разговоров и переписки, и пр. И государство, и человек обязаны воздерживаться от действий в сфере информации, которые нанесут вред другому лицу. Здесь право на информацию предотвращает нежелательное вмешательство в свободу личности и представляет собой основу индивидуальных свобод, обеспечивающих независимость от произвольного вмешательства в частную жизнь.

Заключение

Обобщение всего вышеприведенного позволяет заключить, что содержание права на информацию выявляют его основные свойства. В контексте информации открытого доступа право на информацию представляет собой одну из фундаментальных возможностей человека, необходимую для его нормального существования, развития и самореализации в политической, экономической и культурной жизни общества. Доступ к открытой информации можно охарактеризовать как право: приоритетное, универсальное, гарантируемое и охраняемое публичной властью.

Развитие международного и национального законодательства, юридической доктрины способствует постепенному расширению спектра прав и свобод человека и гражданина. Поэтому целесообразно говорить об информационных правах и свободах человека и гражданина во всех сферах общественных отношений, которые объединены в рамках информационных прав и свобод.

В свою очередь, реализация права на информацию в контексте правового режима информации ограниченного доступа – осуществляется через применение ограничительных, запрещающих и принудительных правовых средств в рамках специальных административно-правовых режимов. Их целью являются устранение существующей конкуренции между интересами определенных субъектов права, а также защита прав человека, интересов общества и государства.

Литература

1. Бойко Е.В. Обязательства государства по ст. 2 Конвенции о защите прав человека и основных свобод 1950 г. // Актуальные проблемы российского права. 2009. № 2. С. 335-340.
2. Ильин И.А. Общее учение о праве и государстве (фрагменты) // Правоведение. 1992. № 3. С. 93-99.
3. Малько А. Право гражданина на информацию // Общественные науки и современность. 1995. № 5. С. 56-62.
4. Огородов Д.В. Правовое регулирование отношений в информационной сфере: автореф. дис...канд. юр. наук.: спец. 12.00.03 / Российская академия наук, Институт государства и права. М., 2002. 21 с.

CONTENT OF THE RIGHT TO INFORMATION IN THE CONTEXT OF OPEN ACCESS AND RESTRICTED ACCESS INFORMATION MODE

Lyaskovsky Ivan Ivanovich

Phd degree applicant in legal sciences

State Customs Committee of the Republic of Belarus, Vitebsk customs, chief inspector of the legal department

Vitebsk, Republic of Belarus

ivanl@inbox.ru

Abstract

The right to information is presented as an ordered system of competences, as well as restrictions and prohibitions, which are used in connection with the provision of the balance of interests of man, society and the state as participants in information relations. Within the framework of the legal regime of open information, it is considered active and passive forms of exercising the right to information. Within the framework of the legal regime of restricted information – some legal values, which founds the possibility of restriction of the right to information.

Keywords

legal regimes of information; open information; restricted information; legislation of the Republic of Belarus; information law; right to information

References

1. Boyko Ye.V. Obyazatel'stva gosudarstva po st. 2 Konventsii o zashchite prav cheloveka i osnovnykh svobod 1950 g. // Aktual'nyye problemy rossiyskogo prava. 2009. № 2. S. 335-340.
2. Il'in I.A. Obshcheye ucheniye o prave i gosudarstve (fragmenty) // Pravovedeniye. 1992. № 3. S. 93-99.
3. Mal'ko A. Pravo grazhdanina na informatsiyu // Obshchestvennyye nauki i sovremennost'. 1995. № 5. S. 56-62.
4. Ogorodov D.V. Pravovoye regulirovaniye otnosheniy v informatsionnoy sfere: avtoref. dis...kand. yur. nauk.: spets. 12.00.03 / Rossiyskaya akademiya nauk, Institut gosudarstva i prava. M., 2002. 21 s.

Доверие и безопасность в информационном обществе

ОСОБЕННОСТИ «ЧЕРНОГО РЫНКА» ПЕРСОНАЛЬНЫХ ДАННЫХ И СОЗДАВАЕМЫЕ ИМ ПРОБЛЕМЫ

Статья рекомендована к публикации председателем редакционного совета Ю.Е. Хохловым 28.12.2020.

Алигулиев Расим Магамед оглу

*Доктор технических наук, профессор
Институт информационных технологий НАНА
Баку, Азербайджанская Республика
r.alguliev@gmail.com*

Махмудов Расим Шариф оглу

*Институт информационных технологий НАНА
Баку, Азербайджанская Республика
rasimmahmudov@gmail.com*

Аннотация

В статье исследуются сущность и особенности «черного рынка» персональных данных. Рассмотрены являющиеся предметом купли-продажи на «черном рынке» виды информации и услуг, технологии и их ценовая политика, а также цели, для которых используются персональные данные, приобретенные на «черном рынке». Кроме того, отмечены проблемы, связанные с незаконным приобретением и продажей персональных данных с точки зрения личных, корпоративных и национальных интересов.

Ключевые слова

личная жизнь, персональные данные; «черный рынок» персональных данных; стоимость персональных данных; кража персональных данных

Введение

Базы данных, которые из-за их стратегической важности называют «валютой XXI века», по своей значимости сравнивают с запасами нефти. Развитие в мире инновационных инициатив, информационной экономики и экономики знаний также резко увеличивает спрос на персональные данные. В «Отчете о цифровой экономике» UNCTAD (United Nations Conference on Trade and Development), представленном в 2019 году, говорится: «Основу цифровой экономики составляют цифровые данные. А основу цифровых данных – персональные данные.» [1].

Очевидно, что по мере роста ценности и важности персональных данных растет и спрос на них. Из-за наличия норм юридической ответственности и ограничений на использование персональных данных спрос на их использование в незаконных деловых и криминальных целях способствует формированию «черного рынка» в этой сфере.

Применение современных технологий и инструментов информационной безопасности в основном обеспечивают защиту от стандартных кибератак. Поэтому киберпреступники постоянно ищут новые возможности и методы обхода существующих систем безопасности. Одним из таких методов является получение персональных данных посредством «черного рынка» [2].

На «черном рынке» могут быть проданы любые персональные данные, которые люди хотят сохранить в секрете от других, например, удостоверение личности, водительские права, страховое свидетельство, кредитная карта, карта здоровья, банковский счет, ИНН, информация о мобильном телефоне, адрес электронной почты, домашний или рабочий адрес и т. д. Киберпреступники используют эту информацию для мошенничества, шантажа и вымогательства, применяя методы социальной инженерии. Хотя исследователи предсказывали формирование рынка персональных данных еще в 1990-х годах, представители научного и юридического сообществ в то время относились к этой идее весьма осторожно. Они считали, что право на неприкосновенность частной

жизни и существующие механизмы для его реализации не позволят персональным данным стать предметом торговли [3].

Многие эксперты по-прежнему считают неприкосновенность частной жизни неотъемлемым правом человека. Хотя технологии конфиденциальности нацелены на защиту этих прав, рынок персональных данных развивается в другом направлении. Многие эксперты предсказывают, что «черный рынок» персональных данных будет продолжать расти. Этот быстрорастущий «черный бизнес» поддерживается мировым сообществом хакеров и разработчиков вредоносных программ. Таким образом, «черный рынок» персональных данных также стимулирует развитие рынка вредоносного и криминального программного обеспечения [4].

В Соединенных Штатах к товарам «черного рынка» проявляют большой интерес. В 2019 году страна потратила на персональные данные 15,2 миллиарда долларов. Самый крупный «черный рынок» персональных данных в мире находится в США – около 60% мирового рынка. Китайский рынок персональных данных с 2,4 миллиардами долларов является вторым по величине в мире. Следующие три места занимают Великобритания, Канада и Франция соответственно [5].

В статье исследуются сущность и особенности «черного рынка» персональных данных. Рассмотрены виды информации и услуг, технологии и их ценовая политика, которые являются предметом торговли на «черном рынке», определена цель использования персональных данных, приобретенных на «черном рынке». Кроме того, отмечены проблемы, связанные с незаконным приобретением и продажей этих данных с точки зрения личных, корпоративных и национальных интересов.

1. Подходы к сущности личной жизни и персональных данных

Персональные данные являются важной составляющей личной жизни. Другими словами, обеспечение конфиденциальности личной жизни требует защиты персональных данных. Поэтому, чтобы лучше понять суть персональных данных, целесообразно рассматривать их через призму личной жизни [6].

Благодаря практическому применению статьи 8 «Европейской конвенции о защите прав и свобод человека» (1950 г.), дано точное нормативное объяснение термина «личная жизнь». Так, в заявлении Европейского суда по правам человека (1992 г.) говорится: «Личная жизнь – это широкая категория, которую невозможно полностью описать. Каждый человек волен развивать эту концепцию и придавать ей определенный смысл. Недопустимо ограничивать это понятие «внутренним кругом» и исключать внешний мир, не принадлежащий этому кругу. Таким образом, понятие «личная жизнь» отражает необходимость развития права на взаимодействие с другими людьми и внешним миром» [7].

Общим эквивалентом термина «неприкосновенность частной жизни» и термина «privacy», используемого в международных документах, является латинское выражение «конфиденциальность». Его смысл – свобода, секретность, одиночество, собственность, личная жизнь, частная жизнь, неприкосновенность, неприкосновенность частной жизни, идентичность, межличностные отношения.

Персональные данные можно разделить на две категории в зависимости от отношения к ним людей, к которым они относятся [2]:

- «нейтральные» персонализированные данные, факт обнаружения и распространения которых субъект этих данных игнорирует;
- данные, распространение которых субъект этих данных хочет ограничить.

Данные, относящиеся ко второй категории, называются «персональными данными», а несанкционированный доступ и использование этих данных классифицируются в международном праве как «нарушение прав субъекта данных».

2. Основные особенности и услуги «черного рынка»

«Черный рынок» традиционно определяется как место незаконного обмена товарами и услугами. Очевидно, что существуют такие виды товаров и услуг, продажа которых разрешается законом. Однако в некоторых случаях этот процесс осуществляется без официального оформления (без получения лицензии, без учета налогов или без соблюдения других необходимых требований). В этом случае покупка и продажа товаров и услуг попадает в категорию «черного рынка». Есть также товары и услуги, продажа которых вообще запрещена законом (например, наркотики, торговля

людьми и т. д.). Закон не разрешает покупку и продажу персональных данных без разрешения владельцев. Следовательно, незаконная покупка и продажа такой информации также попадает в категорию «черного рынка».

Следует отметить, что купля-продажа персональных данных – это новый вид бизнеса, который становится все более актуальным в связи с развитием Интернета и других информационных технологий, электронных сервисов, виртуальных отношений. Персональные данные тогда становятся предметом торговли, когда определенная категория людей заинтересована в их использовании для бизнеса и других целей.

Одна из основных причин широкого использования персональных данных на «черном рынке» заключается в том, что эта сфера пока не регулируется законодательством. Пока что законодательство ни одной страны не признает какую-либо информацию, в том числе персональные данные, предметом продажи. Поэтому такие выражения, как «покупка и продажа персональных данных», «кража персональных данных» юридически описать сложно.

На данный момент только закон «О защите персональных данных потребителей», принятый штатом Калифорния и вступивший в силу 1 января 2020 года, предусматривает куплю-продажу персональных данных [8]. Согласно этому закону, продажа персональных данных – это предоставление компанией персональных данных потребителя в коммерческих целях третьей стороне в устной, письменной, электронной или других формах за плату. В данном случае речь идет о продаже персональных данных, собранных компанией от потребителей, которых она обслуживает, третьему лицу, которое желает использовать эту информацию в коммерческих целях для получения дохода. То есть этот закон не регулирует прямые отношения между владельцами персональных данных и их получателями. Эксперты прогнозируют, что такой правовой механизм в будущем будет применяться на всей территории США, а также в других странах.

Термин «кража персональных данных» как юридическое понятие в настоящее время отражен только в законодательстве США и Великобритании [9]. «Кража персональных данных» означает захват такого типа информации другими лицами различными скрытыми способами в коммерческих, мошеннических или других преступных целях без разрешения самого владельца персональных данных. В законодательствах многих стран такие действия, как кража персональных данных, описывается как «нарушение неприкосновенности личной жизни», «несанкционированный доступ к компьютерной информации» и т. д.

Однако, несмотря на то, что эти юридические вопросы еще не решены, персональные данные как предмет купли-продажи уже давно пользуются большим спросом для целей Big Data. Отметим, что Big Data – это одна из основных технологических и экономических тенденций в мире.

Одна из основных характеристик персональных данных – это их ценность. Ценность персональных данных определяется целью человека, который ее ищет. Как и в случае с другими предметами торговли, рыночный закон спроса и предложения играет важную роль в определении ценности персональных данных.

Цена при продаже персональных данных на «черном рынке» зависит от социального статуса и финансовых возможностей людей. Для формирования кадрового состава этого «черного рынка труда» организуются специальные хакерские курсы. В большинстве случаев оплата на «черном рынке» производится в криптовалюте, что очень затрудняет борьбу с подобной киберпреступностью. На формирование цен персональных данных на черном рынке влияют 4 основных фактора. Во-первых, как и в нормальной экономике, на «черном рынке» действуют законы спроса и предложения. Вторым фактором является остаток на счете – если на кредитной карте доступная сумма в достаточном количестве, то и цена будет высокой. Третий фактор – это высокий бал стабильности карт лояльности (loyalty cards), что также увеличивает их ценность. Наконец, ценность персональных данных может зависеть от их повторного использования. То есть возможность повторного использования увеличивает ценность этих данных. Например, клиенты «черного рынка» платят больше за многократную кредитную карту, чем за одноразовую подарочную карту.

Незаконная продажа персональных данных часто осуществляется в Интернете. Для оказания подобных незаконных услуг широко используется скрытая сеть DarkNet. Доступ к этой сети возможен с помощью специального программного обеспечения. Анонимность членов этой сети защищена, и они обязаны осуществлять куплю-продажу посредством биткойна и других криптовалют [10].

На «черном рынке» корпоративные данные обычно более ценны, чем персональные. Поэтому преступники больше нацелены на персональные данные, которые собираются и хранятся на предприятиях и в организациях, что делает их более уязвимыми.

Киберпреступники сортируют и инвентаризируют любую незаконно изъятую информацию без разрешения владельца, прежде чем выпустить ее на рынок. Данным, которые считают ценными, они назначают более высокую цену. А то, что, по их мнению, менее ценно, они собирают и массово продают за небольшую плату.

Хакеры и посредники с «черного рынка» пытаются по возможности быстро провести процесс кушпи-продажи украденных персональных данных, потому что их владельцы стараются принять ответные меры, как только узнают о краже.

Если какие-либо персональные данные являются актуальными, надежными и полными, то они считаются ценным ресурсом для бизнеса. Персональные данные, обладающие всеми этими свойствами, пользуются большим спросом на «черном рынке». Однако, поскольку такого рода информация приобретает нелегальным путем, проверить ее невозможно.

На «черном рынке» персональных данных предлагаются следующие услуги [10]:

- местонахождение мобильного абонента в режиме реального времени;
- идентификация лиц, находящихся в зоне охвата мобильного абонента в режиме реального времени;
- список лиц, с которыми абонент общается по мобильному телефону и другая информация;
- информация о географическом местоположении, времени и траекториях движения человека;
- информация о личном имуществе человека, его банковских счетах и т. д.;
- персональные данные, собираемые на серверах провайдеров при использовании интернет-сервисов и т.д.

3. Почему персональные данные покупают на «черном рынке»?

На «черном рынке» персональные данные покупают для различных видов деятельности. Но в большинстве случаев конечная цель одна – получить прибыль. Есть разные способы незаконного заработка денег с помощью персональных данных. Рост количества компаний, специализирующихся на аналитике Big Data, разработке программного обеспечения для автоматизации маркетинговых процессов, быстро увеличивает спрос на персональные данные на «черном рынке». В цифровом маркетинге существует наибольший спрос на информацию, которая служит следующим целям [11]:

- повышение эффективности интернет-компаний;
- точный таргетинг рекламных кампаний;
- анализ профилей и интересов пользователей;
- разработка наилучших продуктов и услуг.

На основе анализа полученных персональных данных коммерческие компании определяют целевую аудиторию, применяют индивидуализированные, адресные рекламные технологии. Таким образом, когда люди используют различные поисковые системы, социальные сети, онлайн-сервисы в Интернете, собирается информация об их желаниях, потребностях, образе жизни, увлечениях, которая становится очень полезной для рекламной и маркетинговой деятельности.

Одно из самых распространенных преступлений -- это кража финансовой информации чужих кредитных карт, банковских счетов и т. д. и присвоение их денег. Для получения такой информации используются различные технологии, характерные для физической и виртуальной среды.

В некоторых случаях персональные данные необходимы для создания и продажи поддельных документов. В других случаях – вымогают деньги, шантажируя людей после изъятия их персональных данных. Кроме того, криминальные группы, получив информацию о наличии у кого-то ценного имущества, пытаются заполучить его. Персональные данные также могут быть предметом политической борьбы. Так, политические оппоненты получив незаконным путем персональные данные другой стороны, могут использовать их для снижения ее политического престижа.

Персональные данные также могут использоваться иностранными спецслужбами в политических и идеологических целях. В этом случае персональные данные становятся объектом национальной безопасности.

4. Проблемы, создаваемые «черным рынком»

Формирование и развитие «черного рынка» персональных данных создают ряд проблем для государства, граждан, предприятий и организаций. Во-первых, персональные данные, попадающие на «черный рынок», приобретаются незаконным путем, преступными методами, с причинением вреда юридическим и физическим лицам. Во-вторых, во многих случаях персональные данные используются в злонамеренных целях.

Кража персональных данных может иметь ряд негативных последствий для их владельцев. Согласно опросу, 40% людей, чьи персональные данные были украдены, заявили, что не могут нормально спать по ночам, 65% испытывали чувство нервозности, 69% – чувство опасности и страха. 7% респондентов заявили, что почти дошли до самоубийства, 15% жертв кражи продали свои личные вещи для покрытия убытков, а 7% взяли для из-за этого ссуду [12].

Кража персональных данных наносит большой ущерб бизнесу. Факты показывают, что это постоянный риск для компаний и к подобным негативным ситуациям всегда нужно быть готовым. Для компаний, столкнувшихся с утечками данных, самая большая финансовая потеря – это потеря репутации. После таких случаев компании должны принять необходимые меры для восстановления доверия клиентов и снижения долгосрочных отрицательных финансовых последствий.

Большинство утечек данных – это результат кибератак. На расследование и выявление подобных инцидентов уходит много времени. Компании признают, что чем больше времени требуется на обнаружение и предотвращение утечек данных, тем дороже это стоит. Объем таких расходов растет с каждым годом. Для выявления и предотвращения подобных угроз необходимо использовать личный опыт и увеличивать инвестиции в технологии [4].

В усиленно регулируемых областях, таких как здравоохранение и финансовые услуги, утечка данных обходится дороже, потому что уровень и вероятность потери репутации и клиентов в результате подобных инцидентов в этих сферах выше.

Если персональные данные попадут в руки внешней разведки через «черный рынок», то могут возникнуть серьезные проблемы для национальной безопасности. Так, с помощью технологий Big Data можно проанализировать большой объем персональных данных, из которых можно извлечь различную полезную информацию, интересующую иностранные спецслужбы. Эта информация может отражать отношение, взгляды, подходы граждан к любому вопросу в любой стране или их материальное, моральное, психологическое состояние и состояние здоровья. Кроме того, эта информация может быть использована спецслужбами для совершения любых провокаций против страны.

Другой преступной целью получения персональных данных через «черный рынок» является использование их для вымогательства посредством шантажа и запугивания. В этом случае шантажируемые чиновники вынуждены подчиняться требованиям шантажиста из опасения нанести ущерб своей карьере и распаду семейных отношений [6].

Заключение

Для эффективной борьбы с «черным рынком» персональных данных важно усовершенствовать соответствующее законодательство и усилить меры по борьбе с киберпреступностью.

С технологической точки зрения ответственность за решение важных задач ложится на учреждения, реализующие государственную политику по усилению защиты персональных данных, а также предприятия и организации, оказывающие различные услуги гражданам.

Сами граждане должны заботиться о своих персональных данных, правильно обращаться с ними, учитывать особенности виртуальной среды. Особое внимание следует уделить формированию у них культуры информационной безопасности.

Кроме того, чтобы эффективно использовать персональные данные для нужд Big Data, необходимо разработать и внедрить механизмы их анонимности.

В настоящее время функционирующие в этой сфере международные юридические органы обсуждают создание соответствующих юридических механизмов, позволяющих людям пользоваться своими персональными данными, получая при этом доход. Для этого, прежде всего, важно, чтобы каждый получил право на собственность своих персональных данных.

Незаконный доступ к персональным данным может стать серьезной угрозой национальной безопасности. Таким образом, защиту персональных данных следует рассматривать как неотъемлемую часть национальной безопасности.

Следует отметить, что большие данные, в которых персональные данные имеют особый вес, сравниваются с запасами нефти с точки зрения их стратегической важности и преимуществ. Следовательно, необходимо сделать так, чтобы и права людей на неприкосновенность частной жизни, и национальная безопасность были обеспечены на высоком уровне, а также чтобы была возможность использовать персональные данные в полезных целях.

Литература

1. UNCTAD. Digital Economy Report (2019), United Nations, 2019, 194 p.
2. Alguliyev R., Mahmudov R. Sensitivity of personal data in the context of national mentality and issues of ensuring their security // Problems of Information Society, 2019, № 2, pp. 117-128.
3. Samuelson P. Privacy as intellectual property? // Stanford Law Review, 2000, pp. 1125-1173.
4. Nuncic M. The Black Market for Data. <https://www.ontrack.com> (дата обращения: 15.09.2020).
5. Fontinelle A. How Black Markets Work. 2019. <https://www.investopedia.com> (дата обращения: 23.09.2020).
6. Alguliyev R., Mahmudov R., The place of personal data in the national mentality and the problems of their protection / V Republican Conference "Actual multidisciplinary scientific-practical problems of information security", Baku, November 29, 2019, pp. 21-24.
7. Красотенко О. Понятие «частная жизнь» в решениях Европейского Суда по правам человека / Сборник тезисов 68-й научно-практической конференции студентов, магистрантов и аспирантов факультета международных отношений БГУ. Минск, 27 апреля 2011, с. 51-53.
8. Dayman D. CCPA Sell Definition, 29 September, 2020. <https://www.osano.com/articles/ccpa-definition-sell> (дата обращения: 18.09.2020).
9. Oxford English Dictionary. <https://www.oxfordlearnersdictionaries.com> (дата обращения: 10.11.2020).
10. Armor. The Black Market Report, 2018, 16 p.
11. Spiekermann S., Böhme R., The challenges of personal data markets and privacy // Electronic Markets, June 2015, <https://www.eprofing.springer.com/journals/printpage> (дата обращения: 06.10.2020).
12. Cook G. How Much Is My Identity Worth on the Black Market? <https://www.findreviews.com> (дата обращения: 08.10.2020).

FEATURES OF THE “BLACK MARKET” OF PERSONAL DATA AND THE PROBLEMS CREATED BY THEM

Aliguliyev, Rasim M.

*Doctor of technical sciences, professor
Institute of Information Technologies of ANAS
Baku, Azerbaijan
r.alguliev@gmail.com*

Mahmudov, Rasim Sh.

*Institute of Information Technologies of ANAS
Baku, Azerbaijan
rasimmahmudov@gmail.com*

Abstract

The article examines the essence and features of the “black market” of personal data. The types of information and services, technologies and their pricing policy, as well as the purposes for which personal data purchased on the black market are used, are considered. In addition, the problems associated with the illegal acquisition and sale of personal data from the point of view of personal, corporate and national interests were noted.

Keywords

personal life, personal data; “black market” of personal data; the cost of personal data; theft of personal data

References

1. UNCTAD. Digital Economy Report (2019), United Nations, 2019, 194 p.
2. Alguliyev R., Mahmudov R. Sensitivity of personal data in the context of national mentality and issues of ensuring their security // Problems of Information Society, 2019, № 2, pp. 117-128.
3. Samuelson P. Privacy as intellectual property? // Stanford Law Review, 2000, pp. 1125-1173.
4. Nuncic M. The Black Market for Data. <https://www.ontrack.com> (accessed: 15.09.2020).
5. Fontinelle A. How Black Markets Work. 2019, <https://www.investopedia.com> (accessed: 23.09.2020).
6. Alguliyev R., Mahmudov R., The place of personal data in the national mentality and the problems of their protection / V Republican Conference “Actual multidisciplinary scientific-practical problems of information security”, Baku, November 29, 2019, pp. 21-24.
7. Krasotenko O. Ponyatiye “chastnaya zhizn” v resheniyakh Evropeyskogo Suda po pravam cheloveka / Sbornik tezisov 68-y nauchno-prakticheskoy konferentsii studentov, magistrantov i aspirantov fakulteta mezhdunarodnykh otnosheniy BGU. Minsk. 27 aprelya 2011. s. 51-53.
8. Dayman D. CCPA Sell Definition, 29 September, 2020, <https://www.osano.com/articles/ccpa-definition-sell> (accessed: 18.09.2020).
9. Oxford English Dictionary, <https://www.oxfordlearnersdictionaries.com> (accessed: 10.11.2020).
10. Armor. The Black Market Report, 2018, 16 p.
11. Spiekermann S., Böhme R., The challenges of personal data markets and privacy // Electronic Markets, June 2015, <https://www.eprofing.springer.com/journals/printpage> (accessed: 06.10.2020).
12. Cook G. How Much Is My Identity Worth on the Black Market? <https://www.findreviews.com> (accessed: 08.10.2020).

Технологии информационного общества

ПРОГНОЗИРОВАНИЕ БАНКРОТСТВА МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

Статья рекомендована к публикации членом редакционного совета А.Н. Райковым 23.11.2020.

Доржиев Ардан Саянович

*Финансовый университет при Правительстве Российской Федерации, факультет информационных технологий и анализа больших данных, магистрант
Москва, Россия
a.s.dorzhev@mail.ru*

Аннотация

В своевременном прогнозировании банкротства заинтересованы многие стороны: акционеры, руководители и сотрудники самой компании, контрагенты, партнеры, кредиторы и, конечно, регуляторы и органы государственной власти. В статье представлены существующие модели прогнозирования банкротства: как традиционные статистические, так и современные, на основе машинного обучения, описаны ограничения в применении моделей, способы работы с несбалансированными данными, а также показано, в чем заключается превосходство современных методов прогнозирования над традиционными статистическими.

Ключевые слова

прогнозирование банкротства, Z-оценка, дискриминантный анализ, машинное обучение, нейронные сети, случайные леса, метод опорных векторов

Введение

Прогнозирование финансовой неустойчивости – важный компонент управления компанией. Из-за банкротства¹ акционеры теряют доходы, бизнес, нарушаются цепочки поставок, фискальные органы лишаются налоговых поступлений, госорганы фиксируют снижение экономического роста и повышение социальной напряженности, а работники вынуждены искать новую работу. Поэтому получение ясной картины финансового и имущественного состояния компаний уже много лет является целью специалистов самых разных областей знаний.

Компании активно внедряют современные технологии в основную деятельность, автоматизируя большое количество бизнес-процессов [1]. Это позволяет выстраивать, например, электронную экспертизу, с помощью которой возможно оперативно объединять множество экспертов из разных предметных областей для полного охвата рассматриваемой проблемы и принятия коллективного решения [41]. Электронная экспертиза также подразумевает взаимодействие людей с интеллектуальными системами, которые способны строить анализ и прогнозы на основе более широкого пространства переменных. Так, в частности, методы машинного обучения применяются уже в значительном количестве бизнес-задач [71], в том числе для интеллектуальной аналитики больших данных [14], которые компания аккумулирует, и для построения моделей прогнозирования.

Большинство таких моделей, которые будут рассмотрены далее, относятся к классу инерционных, то есть тех, которые предоставляют прогноз, основанный на продолжении некоторого тренда из предыдущих временных периодов. Однако риски банкротства могут также возникнуть из-за некорректного стратегического менеджмента. Такой менеджмент подразумевает разработку

¹ Для краткости в данной статье термин «банкротство» употребляется для обозначения состояния, при котором должник не способен рассчитаться со своими кредиторами

© Доржиев А.С., 2021. Производство и хостинг журнала «Информационное общество» осуществляется Институтом развития информационного общества.

Данная статья распространяется на условиях международной лицензии Creative Commons «Атрибуция — Некоммерческое использование — На тех же условиях» Всемирная 4.0 (Creative Commons Attribution – NonCommercial – ShareAlike 4.0 International; CC BY-NC-SA 4.0). См. <https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode.ru>

долгосрочных целей и действий, которые позволят достичь более высоких результатов в будущем, например, стать лидирующей компанией в своей отрасли. Разрабатываемые при этом стратегии обычно носят амбициозный характер, поэтому цели компании в таком случае не направлены на пролонгацию сложившейся динамики. Правильный анализ стратегической ситуации также важен для прогнозирования банкротства. Существует множество методов для ее оценки, в том числе и на основе анализа больших данных [67]. Но это отдельная категория прогнозов, и в рамках данной работы рассмотрена не будет.

Данная работа посвящена различным методам машинного обучения для решения задачи прогнозирования банкротства и их сравнению с традиционными методами статистики.

Обзор методов прогнозирования

Один из первых подходов к анализу вероятности банкротства была Z-оценка Альтмана. В своем исследовании [29] он использовал мультивариативный дискриминантный анализ для изучения 66 компаний. Его Z-оценка представляет собой линейную комбинацию следующих показателей компании:

1. Текущие активы / общие активы.
2. Нераспределённая прибыль / общие активы.
3. Прибыль до уплаты налогов / общие активы.
4. Рыночная стоимость компании / общая сумма долга.
5. Продажи/ общие активы.

Альтман вывел формулу, которая учитывает показатели компании и выдает оценку, по которой определяется финансовое состояние компании, и то, насколько вероятно, что она станет банкротом.

В 1980 году американский специалист Ольсон стал одним из первых, кто применил метод логистической регрессии к оценке вероятности дефолта компании [61]. Его модель основана на анализе большего числа признаков, чем модель Альтмана:

1. Натуральный логарифм отношения совокупных активов к индексу-дефлятору валового внутреннего продукта (ВВП).
2. Совокупные обязательства.
3. Совокупные активы.
4. Рабочий капитал.
5. Краткосрочные обязательства.
6. Оборотные активы.
7. Чистая прибыль.
8. Чистая прибыль + Амортизация.

Модель Ольсона дает более интерпретируемый результат, который находится в промежутке между 0 и 1 и является вероятностью наступления дефолта. В дискриминантных моделях вероятность банкротства не имеет номинального значения. Поэтому, в логистических, в отличие от дискриминантных, отсутствуют «серые» зоны, то есть такие зоны, при попадании в которые невозможно однозначно сделать вывод о вероятности банкротства.

Дискриминантный анализ и логистическая регрессия использовались в дальнейшем широко [43] [73], но при этом прогнозные модели строились на относительно малых объемах данных [44]. Современные вычислительные мощности позволяют строить модели на анализе больших данных, используя машинное обучение [76].

Процесс создания прогнозных моделей на основе машинного обучения начинается с подготовки данных. Далее выбираются метрики для оценки результатов, подбираются алгоритмы и определяются их гиперпараметры, с помощью которых оптимизируется их работа. Этап работы с данными состоит не только из их очистки, при котором из выборки удаляются выбросы - значения, сильно выбивающиеся из общего тренда или распределения, но также здесь проводят нормализацию и стандартизацию для численных данных, кодирование для категориальных данных, например, перевод значений “М” и “Ж” признака “Пол” в числа 0 и 1 [31]. Это необходимо, так как большинство алгоритмов машинного обучения способно работать только с численными данными. Дополнительно из данных можно получать метаданные путем поиска комбинаций исходных в выборке. Процесс создания новых признаков из уже имеющихся называется feature engineering [39].

Когда данные обработаны, их разделяют на две группы: целевой признак, который содержит ответы, например, банкрот/не банкрот, и признаки, которые будут обрабатываться алгоритмом машинного обучения для прогнозирования целевого признака [56].

Для запуска, обучения и тестирования моделей всю выборку делят на несколько частей: тренировочные, проверочные и тестовые. С помощью первой обучают модель, вторая необходима для корректировки и поиска наилучших параметров алгоритма, а с помощью последней, которая представляет собой данные, которые модель ни разу не видела, оценивают работу построенной модели. Далее проводится поиск оптимального алгоритма и определение таких его параметров, которые позволяют получить наиболее точный прогноз.

Среди последних исследований по теме прогнозирования банкротства много работ посвящено сравнению стандартных моделей, основанных на методах статистики, и современных моделей на основе машинного обучения [6][32]. Так, например, в работе [32, с. 405-417] авторы исследовали показатели американских и канадских компаний с 1985 по 2013. Результаты исследования показали, что точность прогнозов, полученных с помощью машинного обучения в среднем выше на 10-20%, чем у моделей на основе дискриминантного анализа и логистической регрессии. При этом авторы отмечают, что для построения моделей они не проводили специальную работу над данными: стандартизацию или создание новых признаков, так как в рамках данного исследования им было важно изучить возможность применения алгоритмов обучения без обработки данных. Однако они подчеркивают, что тщательная работа с признаками – это самый перспективный способ улучшения прогнозных моделей на основе машинного обучения.

Один из вариантов работы с признаками был предложен в статье [53, с. 161-188], где австралийские специалисты Стюарт Джонс и Тим Ванг представили новый алгоритм машинного обучения «TreeNet». Имея в распоряжении данные по 4,5 миллионам компаний, они построили 3 модели: одну бинарную и два классификатора, которые определяют несколько состояний рассматриваемой компании: действующая, на стадии банкротства, обанкротившаяся, прекратившая свою деятельность не по причине банкротства. К стандартным показателям компании авторы дополнительно добавили внешние факторы риска: макроэкономические параметры (рост безработицы, инфляцию и т.д.), а также другие – нефинансовые показатели компаний (капитал компании на 1 сотрудника, доход на 1 сотрудника, период погашения дебиторской задолженности и т.д.). В результате, добавление новых признаков позволило существенно увеличить точность построенных прогнозных моделей.

Помимо расширения пространства признаков путем включения в модели дополнительных данных существует другой подход к работе с признаками, который был представлен в статье [37] и называется «label proportions». Он заключается в том, что данные приводятся в виде групп, для каждой из которых вычисляется доля каждого класса, например, банкрот или не банкрот. Таким образом, целью обучения является предсказать класс для каждой группы.

В рамках исследований по машинному обучению специалисты строят множество моделей, основанных на разных алгоритмах, чтобы понять, какой из них лучше подойдет для решения поставленной задачи. Наиболее популярными для задачи прогнозирования банкротства являются нейронные сети [28] [35] [45], случайный лес [6] [25] [32], метод опорных векторов [51] [64] [69], а также ансамбли алгоритмов [60] [72].

Приведенные алгоритмы могут быть полезны не только для создания прогнозных моделей, но и для обнаружения скрытых нелинейных связей между признаками, а также для построения более сложных алгоритмов. Так, в работе [28, с.110-122] приведено сравнение моделей на основе алгоритма AdaBoost, который представляет собой группу простых классификаторов, и нейронных сетей. В исследовании использовались данные европейских компаний, среди которых в качестве основных авторы выделили размер компании, ее финансовые показатели, структуру собственности и вид деятельности. В результате было показано, что с помощью алгоритма AdaBoost величина ошибки прогноза снижается почти на 30%.

Несмотря на широкое применение технологий анализа больших данных и машинного обучения существуют некоторые ограничения и сложности в создании интеллектуальных прогнозных моделей. Одной из проблем является несбалансированность данных в исследуемой выборке. Причина заключается в том, что компаний-банкротов намного меньше, чем финансово устойчивых [11]. Дисбаланс классов может привести к искажению метрик алгоритмов и неправильной интерпретации результатов исследования. В работе [11] описаны 2 способа сбалансировать классы в наборе данных. Первый заключается в сокращении большего по численности наблюдений класса и называется

“андерсемплингом”. Второй, соответственно “оверсемплинг”, предполагает копирование наблюдений миноритарного класса. Так, во многих исследованиях авторы обычно прибегают к одной из приведенных методик балансировки классов [6] [11] [30]. Однако, специалистами из Словакии был предложен еще один способ работы с диспропорцией данных. Они предложили модель с использованием метода одноклассовой классификации [77]. В этом случае обучение идет только на том классе, данных по которому больше, а аномалии, получаемые в результате работы модели, помечаются как другой класс.

Другим важным ограничением в применении и внедрении моделей машинного обучения является сложность алгоритмов, на которых они построены. Для не специалистов в области математики и науки о данных большинство алгоритмов машинного обучения – это черный ящик, которому на вход дают данные, а на выходе получают результат. Такие модели плохо интерпретируемы, и бизнес-заказчикам трудно полностью доверить им принятие решений, так как не до конца понятно, как они работают [62, с. 464–473].

Также, большинство прогнозных моделей, которые разработаны на основе данных бухгалтерского анализа, основаны на допущении, что публикуемая финансовая отчетность является полностью достоверной и справедливой. Но в действительности некоторые компании могут специально исказить истинную картину своего финансового состояния в целях сохранения инвестиционной привлекательности или изменения восприятия риска для партнеров. Так, в работе [50] предлагается способ анализа степени влияния таких манипуляций с данными на модели прогнозирования. Для решения этой задачи авторы дополнили модели новым признаком - мерой искажения финансовой отчетности. Эта мера основывается на показателе доходности активов компании. В результате, было показано, что классификаторы с этим новым признаком, который учитывает возможные искажения отчетности компании, дают более точный прогноз, чем те, которые основаны только на чистых данных.

Большинство работ по этой теме сосредоточены на оптимизации старых или создании новых методов достижения наибольшей точности прогнозных моделей. Однако, важно также учитывать и оценивать издержки, связанные с ошибками моделей прогнозирования. В исследовании [38, с. 612-630] представлена методология, основанная на гетерогенном выборе ансамбля алгоритмов и многоцелевой оптимизации затрат из-за ошибочной работы классификатора. Авторы проанализировали компании Франции, Бельгии, Италии и поделили их на группы по 7 отраслям для тестирования предложенного решения. Построив ансамбли классификаторов, отобранных с помощью методов многокритериальной оптимизации, специалисты добились уменьшения издержек от ошибок работы моделей.

Прогнозные модели применительно к российскому рынку

В России модели прогнозирования банкротства начинают развиваться только в 1990-е годы [6]. Первые исследования представляли собой попытки адаптировать к российскому рынку зарубежные модели, которые строились на анализе американских, канадских и европейских компаний [5]. Однако полученные таким образом модели обладали низкой прогнозной силой и требовали существенных дальнейших доработок [10], связанных с более правильным подбором переменных для анализа согласно российскому законодательству о банкротстве и системы бухгалтерского учета нашей страны [26].

При этом в большинстве российских исследований авторы стараются сосредоточить внимание на прогнозировании банкротства в рамках одной отрасли [21] или одного сегмента бизнеса [2], так как полагают, что для разных отраслей и типов компаний следует выбирать для анализа и разные наборы данных, методы и алгоритмы [22].

Также особую значимость при построении прогнозных моделей имеет и обработка внешних данных. Так, в работе [25] авторы проанализировали российские компании строительной, торговой отрасли и отрасли обрабатывающего производства для прогнозирования их финансового состояния. Им удалось выявить важные внешние факторы, а именно: прирост валового внутреннего продукта, индекс потребительских цен, прирост курса доллара США, ставка центрального банка, прирост индекса московской биржи, а также уровень инфляции и безработицы.

В исследовании [6], где представлен анализ компаний малого и среднего бизнеса, основной задачей было сравнение подходов к прогнозированию критического состояния, при котором компания не способна исполнить все свои обязательства и вынуждена стать банкротом. Исследователи проанализировали финансовые показатели и внешние нефинансовые факторы около 1 млн российских компаний из отрасли торговли, строительства, обрабатывающего производства и недвижимости. В результате, построив несколько прогнозных моделей с помощью методов дискриминантного анализа,

логистической регрессии, а также деревьев решений и алгоритма случайного леса, авторы обнаружили сильное влияние нефинансовых факторов таких, как отрасль, федеральный округ, где предприятие зарегистрировано, а также возраст самого предприятия.

Как отмечается в работе [10] российские исследования по данной теме сильно уступают зарубежным. Основным недостатком являются малые выборки исследуемых данных для моделей. В большинстве работ выборки имеют объем менее 1 тысячи наблюдений. Из-за этого становится неясно, как будет вести себя модель на больших данных, и какая у нее будет при этом точность. Также, изучив данные из базы СПАРК-Интерфакс за период с 2006 по 2010 гг., авторы обнаружили значительную долю на российском рынке компаний-однодневок. Добавление таких компаний в выборку для составления модели может негативно повлиять на ее точность. Другими отрицательными факторами, которые присутствуют на российском рынке, являются нелегальная деятельность компаний, случаи криминальных банкротств и плохое качество данных, которое заключается в том, что не все фирмы в полной мере предоставляют финансовые данные [6], а некоторые намеренно искажают их, чтобы извлечь выгоду [10] [50].

Таким образом, построение прогнозных моделей для российских компаний является актуальной задачей, так как не все зарубежные модели удается успешно адаптировать под специфику российской экономики и законодательства.

Заключение

В данной работе были представлены существующие подходы к задаче прогнозирования банкротства компаний, начиная от стандартных методов статистики и заканчивая современными алгоритмами машинного обучения. Как показали приведенные в работе исследования, применение машинного обучения позволило строить модели, предоставляющие более точный прогноз.

Задача прогнозирования является комплексной, поэтому пути решения заключаются не только в выборе подходящего алгоритма. Также важно дополнительно проводить работу с данными, включать в анализ внешние данные, генерировать метаданные путем извлечения ценных артефактов из существующих, также искать новые способы и приемы их обработки.

Компаниям важно правильно оценивать свои показатели, чтобы принимать эффективные управленческие решения. Также важно проверять тех, с кем предстоит сотрудничать, то есть иметь инструмент проверки финансового состояния партнёров, заемщиков, поставщиков и т. д. Поэтому, спрос на подобные модели классификации, прогнозирования, оценки будет только расти [69].

Литература

1. Ариничев И. В., Матвеева Л. Г., Ариничева И. В. Прогнозирование банкротства организации на основе метрических методов интеллектуального анализа данных // Journal of Economic Regulation (Вопросы регулирования экономики). – 2018. – Т. 9. – №. 1.
2. Ариничев И.В., Богдашев И.В. Оценка риска банкротства субъектов малого предпринимательства на основе методов машинного обучения // Вестник РУДН. Серия: Экономика. 2017. №2.
3. Водошнянова Д. В., Уродовских В. Н. Аналитический обзор методов прогнозирования вероятности банкротства предприятий // Концепт. – 2015. – №. 11.
4. Горбатков С. А., Белолищев И. И., Макеева Е. Ю. Выбор системы экономических показателей для диагностики и прогнозирования банкротств на основе нейросетевого байесовского подхода // Финансы: Теория и Практика. – 2013. – №. 4.
5. Демешев Б. Б., Тихонова А. С. Динамика прогнозной силы моделей банкротства для средних и малых российских компаний оптовой и розничной торговли // Корпоративные финансы. – 2014. – Т. 8. – №. 3.
6. Демешев Б. Б., Тихонова А. С. Прогнозирование банкротства российских компаний: межотраслевое сравнение // Экономический журнал Высшей школы экономики. – 2014. – Т. 18. – №. 3.
7. Евстропов М. В. Оценка эффективности моделей прогнозирования банкротства предприятий // Экономический анализ: теория и практика. – 2008. – №. 13.

8. Жеребилова А. Б. Методики предсказания банкротства организации в рамках построения прогнозно-функциональной системы // Государственное и муниципальное управление. Ученые записки СКАГС. – 2018. – №. 2.
9. Зеленков Ю. А., Федорова Е. А. Двух-шаговый метод построения неоднородного ансамбля классификаторов для прогнозирования банкротства.
10. Казаков А. В., Кольшшкин А. В. Разработка моделей прогнозирования банкротства в современных российских условиях // Вестник Санкт-Петербургского университета. Экономика. – 2018. – №. 2.
11. Карминский А. М., Бурехин Р. Н. Сравнительный анализ методов прогнозирования банкротств российских строительных компаний // Бизнес-информатика. – 2019. – Т. 13. – №. 3.
12. Макаров А. М., Гимазетдинов Д. Р., Кавченков Е. А. Выбор модели прогнозирования риска банкротства для российских предприятий, функционирующих в сфере внешнеэкономической деятельности // Вестник Удмуртского университета. Серия «Экономика и право». – 2017. – Т. 27. – №. 1.
13. Макеева Е. Ю., Аршавский И. В. Применение нейронных сетей и семантического анализа для прогнозирования банкротства // Корпоративные финансы. 2014. №4.
14. Матвеевский С. С. Опыт Японии по использованию аналитики больших данных для снижения кредитного риска при финансировании малых и средних предприятий // Вестник университета. – 2019. – №. 10.
15. Никифорова Н. А., Донцова Л. В., Донцов Е. В. Интеллектуальный анализ данных в моделировании финансового состояния предприятий // Финансовый журнал. – 2011. – №. 2.
16. Опекунов А. Н., Кузьмина М. Г. Принципы формирования моделей прогнозирования вероятности банкротства предприятий с использованием элементов машинного обучения // Модели, системы, сети в экономике, технике, природе и обществе. – 2019. – №. 4.
17. Опекунов А. Н., Никитина Н. В., Камардина Н. В. Реализация принципов машинного обучения при построении моделей прогнозирования банкротства предприятий // Проблемы развития предприятий: теория и практика. – 2019. – №. 1-1. – С. 152-156.
18. Синельникова-Мурылева Е. В., Горшкова Т. Г., Макеева Н. В. Прогнозирование дефолтов в российском банковском секторе // Экономическая политика. – 2018. – Т. 13. – №. 2.
19. Уродовских В. Н., Бахаева А. А. Об адекватности моделей оценки риска банкротства отечественных предприятий // Социально-экономические явления и процессы. – 2010. – №. 6.
20. Фёдорова Е. А., Гиленко Е. В., Довженко С. Е. Модели прогнозирования банкротства: особенности российских предприятий // Проблемы прогнозирования. – 2013. – №. 2.
21. Федорова Е. А., Довженко С. Е., Тимофеев Я. В. Какая модель лучше прогнозирует банкротство российских предприятий? // Экономический анализ: теория и практика. – 2014. – №. 41 (392).
22. Фёдорова Е. А., Довженко С. Е., Фёдоров Ф. Ю. Модели прогнозирования банкротства российских предприятий: отраслевые особенности // Проблемы прогнозирования. – 2016. – №. 3 (156).
23. Федорова Е. А., Лазарев М. П., Федин А. В. Прогнозирование банкротства предприятия с учетом факторов внешней среды // Финансовая аналитика: проблемы и решения. – 2016. – №. 42 (324).
24. Федорова Е. А., Мусиенко С. О., Федоров Ф. Ю. Анализ влияния внешних факторов на прогнозирование финансовой несостоятельности российских компаний // Вестник Санкт-Петербургского университета. Экономика. – 2020. – Т. 36. – №. 1.
25. Федорова Е. А., Мусиенко С. О., Федоров Ф. Ю. Прогнозирование банкротства субъектов малого и среднего предпринимательства в России // Финансы и кредит. – 2018. – Т. 24. – №. 11 (779).
26. Федорова Е. А., Чухланцева М. А., Чекризов Д. В. Оценка эффективности прогнозирования банкротства предприятий на основе российского законодательства // Финансы и кредит. – 2017. – Т. 23. – №. 13 (733).
27. Федотова О. В. Прогнозирование банкротства предприятия // Молодой ученый. – 2018. – №. 4. – С. 122-124.
28. Alfaro E., Garcia N., Gamez M., Elizondo D. Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks // Decision Support Systems 45 – 2008. – p. 110-122.
29. Altman E. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy // The Journal of Finance 23 (4) – 1968. – p. 589-609.
30. Alrasheed D., Che D. Improving Bankruptcy Prediction Using Oversampling and Feature Selection Techniques // Proceedings on the International Conference on Artificial Intelligence (ICAI). – The

- Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2018. – p. 440-446.
31. Antunes F., Ribeiro B., Pereira F. Probabilistic modeling and visualization for bankruptcy prediction // *Applied Soft Computing* 60 – 2017. – p. 831-843.
 32. Barboza F., Kimura H., Altman E. Machine learning models and bankruptcy prediction // *Expert Systems with Applications* 83 – 2017. – p. 405-417.
 33. Blanco-Oliver A., Irimia-Dieguez A., Oliver-Alfonso M., Wilson N. Improving Bankruptcy Prediction in Micro-Entities by Using Nonlinear Effects and Non-Financial Variables // *Czech Journal of Economics and Finance* 65 (2) – 2015. – p. 144-165.
 34. Bredart X. Bankruptcy Prediction Model Using Neural Networks // *Accounting and Finance Research* 3, No. 2 – 2014. – p. 124-128.
 35. Burnell P., Folarin B. The Impact of Neural Networks in Finance // *Neural Comput & Applic* 6 – 1997. – p. 193-200.
 36. Chen J. M. Models for Predicting Business Bankruptcies and Their Application to Banking and Financial Regulation // *Penn St. L. Rev.* – 2018. – T. 123. – p. 735.
 37. Chen Z., Chen W., Shi Y. Ensemble learning with label proportions for bankruptcy prediction // *Expert Systems With Applications* 146 – 2020.
 38. De Bock K., Coussement K., Lessmann S. Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach // *European Journal of Operational Research* 285 – 2020. – p. 612-630.
 39. Gogas P., Papadimitriou T., Agrapetidou. Forecasting bank failures and stress testing: A machine learning approach // *International Journal of Forecasting* 34 – 2018. p. 440-455.
 40. Griffin J. M., Lemmon M. L. Book-to-market equity, distress risk, and stock returns // *The Journal of Finance.* – 2002. – T. 57. – №. 5. – p. 2317-2336.
 41. Gubanov D. et al. E-expertise: modern collective intelligence. – Springer International Publishing, 2014.
 42. Hong Hanh L. E., VIVIANI J. L. Predicting Bank Failure: Statistical Technique versus Intelligent Technique.
 43. Hardinata L., Warsito B. Bankruptcy prediction based on financial ratios using Jordan Recurrent Neural Networks: a case study in Polish companies // *Journal of Physics: Conf. Series* 41025 – 2018. – 012098
 44. Hillegeist S. A. et al. Assessing the probability of bankruptcy // *Review of accounting studies.* – 2004. – T. 9. – №. 1. – p. 5-34.
 45. Hosaka T. Bankruptcy prediction using imaged financial ratios and convolutional neural networks // *Expert Systems With Applications* 117 – 2019. – p. 287-299.
 46. Iturriaga F., Sanz. I. Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks // *Expert Systems with Applications* 42 – 2015. - p. 2857-2869.
 47. Jardin P. Bankruptcy prediction and neural networks: The contribution of variable selection methods. // *Proceedings of the Second European Symposium on Time Series Prediction, Helsinki University of Technology, Porvoo, Finland* – 2008. - p. 271-284.
 48. Jardin P. Failure pattern-based ensembles applied to bankruptcy forecasting // *Decision Support Systems* 107 – 2018. – p. 64-77.
 49. Jardin P. Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy // *Neurocomputing* 73, issue 10-12 – 2010. – p. 2047-2060.
 50. Jardin P. Veganzones D., Severin E. Forecasting Corporate Bankruptcy Using Accrual-Based Models // *Comput Econ* 54 – 2019 – p. 7-43.
 51. Jayanthi. J, Suresh Joseph. K, Vaishnavi. J. Bankruptcy Prediction using SVM and Hybrid SVM Survey // *International Journal of Computer Applications* 33(7) – 2011. – p. 39-45.
 52. Jo N.O., Shin K.S. Bankruptcy Prediction Modeling Using Qualitative Information Based on Big Data Analytics // *J Intell Inform Syst.* 22(2) – 2016. – p. 33-56.
 53. Jones S., Wang T. Predicting private company failure: A multi-class analysis // *J. Int. Finance. Markets Inst. Money* 61 – 2019. – p.161-188.
 54. Kasgari A. A. et al. Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses // *Neural Computing and Applications.* – 2013. – T. 23. – №. 3-4. – p. 927-936.
 55. Kasgari A. A. et al. The bankruptcy prediction by neural networks and logistic regression // *International Journal of Academic Research in Accounting, Finance and Management Sciences.* – 2013. – T. 3. – №. 4. – p. 146-152.

56. Kovacova M., Kliestikova J. Modelling bankruptcy prediction models in Slovak companies // SHS Web of Conferences 39 – 2017 – 01013
57. Mai F. et al. Deep learning models for bankruptcy prediction using textual disclosures // European journal of operational research. – 2019. – Т. 274. – №. 2. – С. 743-758.
58. Meese E. N., Viken T. Machine learning in bankruptcy prediction: utilizing machine learning for improved bankruptcy predictions in the Norwegian market with an emphasis on financial, management and sector statements: Master's thesis. – 2019.
59. Nagaraj K., Sridhar A. A predictive system for detection of bankruptcy using machine learning techniques // International Journal of Data Mining & Knowledge Management Process Vol. 5 No. 1. – 2015.
60. Nanni L., Lumini A. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring // Expert Systems With Applications 36 – 2009. – p. 3028–3033.
61. Ohlson J. A. Financial ratios and the probabilistic prediction of bankruptcy // Journal of accounting research. – 1980. – p. 109-131.
62. Olson D., Delen D., Meng Y. Comparative analysis of data mining methods for bankruptcy prediction // Decision Support Systems 52 – 2012. – p. 464–473.
63. Pena T., Martínez S., Abudu B. Bankruptcy prediction: A comparison of some statistical and machine learning techniques // Computational Methods in Economic Dynamics. – Springer, Berlin, Heidelberg, 2011. – p. 109-131.
64. Petropoulos A., Siakoulis V., Stavroulakis V., Vlachogiannakis N. Predicting bank insolvencies using machine learning techniques // International Journal of Forecasting
65. Pompe P.M., Feelders A.J., Using Machine Learning, Neural Networks, and Statistics to Predict Corporate Bankruptcy // Microcomputers in Civil Engineering 12 – 1997. – p. 267–276.
66. Pozorska J., Scherer M. Company Bankruptcy Prediction with Neural Networks // Artificial Intelligence and Soft Computing – 2018. – p.183-189
67. Raikov A. Manufacturer's Strategic Risk Temperature Assessment with Convergent Approach, Cognitive Modelling and Blockchain Technology // IFAC-PapersOnLine. – 2019. – Т. 52. – №. 13. – С. 1289-1294.
68. Shi Y., Li X. An overview of bankruptcy prediction models for corporate firms: A systematic literature review // Intangible Capital 15(2) – 2019.
69. Shin K.S., Lee T.S., Kim H.J. An application of support vector machines in bankruptcy prediction model // Expert Systems with Applications 28 – 2005. – p. 127–135.
70. Soui M., Smiti S., Mkaouer M.W., Ejbali R. Bankruptcy Prediction Using Stacked Auto-Encoders // Applied Artificial Intelligence 2019. – p. 80-100.
71. Suss J., Treitel H. Predicting bank distress in the UK with machine learning // Bank of England Staff Working Paper No. 831 – 2019.
72. Tsai C.F., Wu J.W. Using neural network ensembles for bankruptcy prediction and credit scoring // Expert Systems with Applications 34 – 2008. – p. 2639–2649.
73. Tseng F. M., Lin L. A quadratic interval logit model for forecasting bankruptcy // Omega. – 2005. – Т. 33. – №. 1. – p. 85-91.
74. Wagenmans F. Machine learning in bankruptcy prediction: Master's thesis. – 2017
75. Wang G., Ma J., Yang S. An improved boosting based on feature selection for corporate bankruptcy prediction // Expert Systems with Applications 41 – 2014. – p. 2353–2361.
76. Wang N. Bankruptcy Prediction Using Machine Learning // Journal of Mathematical Finance 7 – 2017. – p. 908-918.
77. Zoricak M., Gnip P., Drotar P., Gazda V. Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets // Economic Modelling 84 – 2020. – p. 165-176.

BANKRUPTCY PREDICTION USING MACHINE LEARNING METHODS

Dorzhiiev Ardan Sayanovich

*Financial University under the Government of the Russian Federation, Department of Information Technologies and Big Data Analysis, Master's degree student
Moscow, Russia
a.s.dorzhiiev@mail.ru*

Abstract

Bankruptcy prediction models are important to many stakeholders starting from associates up to shareholders, creditors and partners. The article contains an overview of both traditional statistic prediction methods and modern ones with the use of machine learning.

Keywords

bankruptcy prediction, Z-score, discriminant analysis, machine learning, neural networks, random forest, support vector machine

References

1. Arinichev I. V., Matveeva L. G., Arinicheva I. V. Prognozirovanie bankrotstva organizacii na osnove metricheskikh metodov intellektual'nogo analiza dannyh // Journal of Economic Regulation (Voprosy regulirovaniya ekonomiki). – 2018. – T. 9. – № 1.
2. Arinichev I.V., Bogdashev I.V. Ocenka riska bankrotstva sub'ektov malogo predprinimatel'stva na osnove metodov mashinnogo obucheniya // Vestnik RUDN. Seriya: Ekonomika. 2017. №2.
3. Vodop'yanova D. V., Urodovskih V. N. Analiticheskij obzor metodov prognozirovaniya veroyatnosti bankrotstva predpriyatij //Koncept. – 2015. – № 11.
4. Gorbatkov S. A., Belolipcev I. I., Makeeva E. YU. Vybor sistemy ekonomicheskikh pokazatelej dlya diagnostiki i prognozirovaniya bankrotstv na osnove nejrosetevogo bajesovskogo podhoda //Finansy: Teoriya i Praktika. – 2013. – № 4.
5. Demeshev B. B., Tihonova A. S. Dinamika prognoznoj sily modelej bankrotstva dlya srednih i malyh rossijskikh kompanij optovoj i roznicnoj trgovli //Korporativnye finansy. – 2014. – T. 8. – № 3.
6. Demeshev B. B., Tihonova A. S. Prognozirovanie bankrotstva rossijskikh kompanij: mezhotraslevoe sravnenie //Ekonomicheskij zhurnal Vysshej shkoly ekonomiki. – 2014. – T. 18. – № 3.
7. Evstropov M. V. Ocenka effektivnosti modelej prognozirovaniya bankrotstva predpriyatij //Ekonomicheskij analiz: teoriya i praktika. – 2008. – № 13.
8. Zherebilova A. B. Metodiki predskazaniya bankrotstva organizacii v ramkah postroeniya prognozno-funktional'noj sistemy //Gosudarstvennoe i municipal'noe upravlenie. Uchenye zapiski SKAGS. – 2018. – № 2.
9. Zelenkov YU. A., Fedorova E. A. Dvuh-shagovyj metod postroeniya neodnorodnogo ansamblya klassifikatorov dlya prognozirovaniya bankrotstva.
10. Kazakov A. V., Kolyshkin A. V. Razrabotka modelej prognozirovaniya bankrotstva v sovremennyh rossijskikh usloviyah //Vestnik Sankt-Peterburgskogo universiteta. Ekonomika. – 2018. – № 2.
11. Karminskij A. M., Burekhin R. N. Sravnitel'nyj analiz metodov prognozirovaniya bankrotstv rossijskikh stroitel'nyh kompanij // Biznes-informatika. – 2019. – T. 13. – № 3.
12. Makarov A. M., Gimazetdinov D. R., Kavchenkov E. A. Vybor modeli prognozirovaniya riska bankrotstva dlya rossijskikh predpriyatij, funkcioniruyushchih v sfere vneshneekonomicheskoy deyatel'nosti //Vestnik Udmurtskogo universiteta. Seriya «Ekonomika i pravo». – 2017. – T. 27. – № 1.
13. Makeeva E. YU., Arshavskij I. V. Primenenie nejronnyh setej i semanticheskogo analiza dlya prognozirovaniya bankrotstva // Korporativnye finansy. 2014. №4.
14. Matveevskij S. S. Opyt YAponii po ispol'zovaniyu analitiki bol'shih dannyh dlya snizheniya kreditnogo riska pri finansirovanii malyh i srednih predpriyatij // Vestnik universiteta. – 2019. – № 10.
15. Nikiforova N. A., Doncova L. V., Doncov E. V. Intellektual'nyj analiz dannyh v modelirovanii finansovogo sostoyaniya predpriyatij // Finansovyj zhurnal. – 2011. – № 2.

16. Opekunov A. N., Kuz'mina M. G. Principy formirovaniya modelej prognozirovaniya veroyatnosti bankrotstva predpriyatij s ispol'zovaniem elementov mashinnogo obucheniya // *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve*. 2019. №4.
17. Opekunov A. N., Nikitina N. V., Kamardina N. V. Realizaciya principov mashinnogo obucheniya pri postroenii modelej prognozirovaniya bankrotstva predpriyatij // *Problemy razvitiya predpriyatij: teoriya i praktika*. – 2019. – №. 1-1. – S. 152-156.
18. Sinel'nikova-Muryleva E. V., Gorshkova T. G., Makeeva N. V. Prognozirovanie defoltov v rossijskom bankovskom sektore // *Ekonomicheskaya politika*. – 2018. – T. 13. – №. 2.
19. Urodovskih V. N., Bahaeva A. A. Ob adekvatnosti modelej ocenki riska bankrotstva otechestvennyh predpriyatij // *Social'no-ekonomicheskie yavleniya i processy*. – 2010. – №. 6.
20. Fyodorova E. A., Gilenko E. V., Dovzhenko S. E. Modeli prognozirovaniya bankrotstva: osobennosti rossijskih predpriyatij // *Problemy prognozirovaniya*. – 2013. – №. 2.
21. Fedorova E. A., Dovzhenko S. E., Timofeev YA. V. Kakaya model' luchshe prognoziruet bankrotstvo rossijskih predpriyatij? // *Ekonomicheskij analiz: teoriya i praktika*. – 2014. – №. 41 (392).
22. Fyodorova E. A., Dovzhenko S. E., Fyodorov F. YU. Modeli prognozirovaniya bankrotstva rossijskih predpriyatij: otraslevye osobennosti // *Problemy prognozirovaniya*. – 2016. – №. 3 (156).
23. Fedorova E. A., Lazarev M. P., Fedin A. V. Prognozirovanie bankrotstva predpriyatija s uchedom faktorov vneshnej sredy // *Finansovaya analitika: problemy i resheniya*. – 2016. – №. 42 (324).
24. Fedorova E. A., Musienko S. O., Fedorov F. YU. Analiz vliyaniya vneshnih faktorov na prognozirovanie finansovoj nesostoyatel'nosti rossijskih kompanij // *Vestnik Sankt-Peterburgskogo universiteta. Ekonomika*. – 2020. – T. 36. – №. 1.
25. Fedorova E. A., Musienko S. O., Fedorov F. YU. Prognozirovanie bankrotstva sub"ektov malogo i srednego predprinimatel'stva v Rossii // *Finansy i kredit*. – 2018. – T. 24. – №. 11 (779).
26. Fedorova E. A., CHuhlanceva M. A., CHekrizov D. V. Ocenka effektivnosti prognozirovaniya bankrotstva predpriyatij na osnove rossijskogo zakonodatel'stva // *Finansy i kredit*. – 2017. – T. 23. – №. 13 (733).
27. Fedotova O. V. Prognozirovanie bankrotstva predpriyatija // *Molodoj uchenyj*. – 2018. – №. 4. – S. 122-124.
28. Alfaro E., Garcia N., Gamez M., Elizondo D. Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks // *Decision Support Systems* 45 – 2008. – p. 110-122.
29. Altman E. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy // *The Journal of Finance* 23 (4) – 1968. – p. 589-609.
30. Alrasheed D., Che D. Improving Bankruptcy Prediction Using Oversampling and Feature Selection Techniques // *Proceedings on the International Conference on Artificial Intelligence (ICAI)*. – The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2018. – p. 440-446.
31. Antunes F., Ribeiro B., Pereira F. Probabilistic modeling and visualization for bankruptcy prediction // *Applied Soft Computing* 60 – 2017. – p. 831-843.
32. Barboza F., Kimura H., Altman E. Machine learning models and bankruptcy prediction // *Expert Systems with Applications* 83 – 2017. – p. 405-417.
33. Blanco-Oliver A., Irimia-Dieguez A., Oliver-Alfonso M., Wilson N. Improving Bankruptcy Prediction in Micro-Entities by Using Nonlinear Effects and Non-Financial Variables // *Czech Journal of Economics and Finance* 65 (2) – 2015. – p. 144-165.
34. Bredart X. Bankruptcy Prediction Model Using Neural Networks // *Accounting and Finance Research* 3, No. 2 – 2014. – p. 124-128.
35. Burnell P., Folarin B. The Impact of Neural Networks in Finance // *Neural Comput & Applic* 6 – 1997. – p. 193-200.
36. Chen J. M. Models for Predicting Business Bankruptcies and Their Application to Banking and Financial Regulation // *Penn St. L. Rev.* – 2018. – T. 123. – p. 735.
37. Chen Z., Chen W., Shi Y. Ensemble learning with label proportions for bankruptcy prediction // *Expert Systems With Applications* 146 – 2020.
38. De Bock K., Coussement K., Lessmann S. Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach // *European Journal of Operational Research* 285 – 2020. – p. 612-630.
39. Gogas P., Papadimitriou T., Agrapetidou. Forecasting bank failures and stress testing: A machine learning approach // *International Journal of Forecasting* 34 – 2018. p. 440-455.

40. Griffin J. M., Lemmon M. L. Book-to-market equity, distress risk, and stock returns // *The Journal of Finance*. – 2002. – Т. 57. – №. 5. – p. 2317-2336.
41. Gubanov D. et al. *E-expertise: modern collective intelligence*. – Springer International Publishing, 2014.
42. Hong Hanh L. E., VIVIANI J. L. Predicting Bank Failure: Statistical Technique versus Intelligent Technique.
43. Hardinata L., Warsito B. Bankruptcy prediction based on financial ratios using Jordan Recurrent Neural Networks: a case study in Polish companies // *Journal of Physics: Conf. Series* 41025 – 2018. – 012098
44. Hillegeist S. A. et al. Assessing the probability of bankruptcy // *Review of accounting studies*. – 2004. – Т. 9. – №. 1. – p. 5-34.
45. Hosaka T. Bankruptcy prediction using imaged financial ratios and convolutional neural networks // *Expert Systems With Applications* 117 – 2019. – p. 287-299.
46. Iturriaga F., Sanz. I. Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks // *Expert Systems with Applications* 42 – 2015. - p. 2857-2869.
47. Jardin P. Bankruptcy prediction and neural networks: The contribution of variable selection methods. // *Proceedings of the Second European Symposium on Time Series Prediction*, Helsinki University of Technology, Porvoo, Finland – 2008. - p. 271-284.
48. Jardin P. Failure pattern-based ensembles applied to bankruptcy forecasting // *Decision Support Systems* 107 – 2018. – p. 64-77.
49. Jardin P. Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy // *Neurocomputing* 73, issue 10-12 – 2010. – p. 2047-2060.
50. Jardin P. Veganzones D., Severin E. Forecasting Corporate Bankruptcy Using Accrual-Based Models // *Comput Econ* 54 – 2019 – p. 7-43.
51. Jayanthi. J, Suresh Joseph. K, Vaishnavi. J. Bankruptcy Prediction using SVM and Hybrid SVM Survey // *International Journal of Computer Applications* 33(7) – 2011. – p. 39-45.
52. Jo N.O., Shin K.S. Bankruptcy Prediction Modeling Using Qualitative Information Based on Big Data Analytics // *J Intell Inform Syst.* 22(2) – 2016. – p. 33-56.
53. Jones S., Wang T. Predicting private company failure: A multi-class analysis // *J. Int. Finance. Markets Inst. Money* 61 – 2019. – p.161-188.
54. Kasgari A. A. et al. Prediction of bankruptcy Iranian corporations through artificial neural network and Probit-based analyses // *Neural Computing and Applications*. – 2013. – Т. 23. – №. 3-4. – p. 927-936.
55. Kasgari A. A. et al. The bankruptcy prediction by neural networks and logistic regression // *International Journal of Academic Research in Accounting, Finance and Management Sciences*. – 2013. – Т. 3. – №. 4. – p. 146-152.
56. Kovacova M., Kliestikova J. Modelling bankruptcy prediction models in Slovak companies // *SHS Web of Conferences* 39 – 2017 – 01013
57. Mai F. et al. Deep learning models for bankruptcy prediction using textual disclosures // *European journal of operational research*. – 2019. – Т. 274. – №. 2. – C. 743-758.
58. Meese E. N., Viken T. Machine learning in bankruptcy prediction: utilizing machine learning for improved bankruptcy predictions in the Norwegian market with an emphasis on financial, management and sector statements: Master's thesis. – 2019.
59. Nagaraj K., Sridhar A. A predictive system for detection of bankruptcy using machine learning techniques // *International Journal of Data Mining & Knowledge Management Process* Vol. 5 No. 1. – 2015.
60. Nanni L., Lumini A. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring // *Expert Systems With Applications* 36 – 2009. – p. 3028-3033.
61. Ohlson J. A. Financial ratios and the probabilistic prediction of bankruptcy // *Journal of accounting research*. – 1980. – p. 109-131.
62. Olson D., Delen D., Meng Y. Comparative analysis of data mining methods for bankruptcy prediction // *Decision Support Systems* 52 – 2012. – p. 464-473.
63. Pena T., Martínez S., Abudu B. Bankruptcy prediction: A comparison of some statistical and machine learning techniques // *Computational Methods in Economic Dynamics*. – Springer, Berlin, Heidelberg, 2011. – p. 109-131.
64. Petropoulos A., Siakoulis V., Stavroulakis V., Vlachogiannakis N. Predicting bank insolvencies using machine learning techniques // *International Journal of Forecasting*

65. Pompe P.M., Feelders A.J., Using Machine Learning, Neural Networks, and Statistics to Predict Corporate Bankruptcy // *Microcomputers in Civil Engineering* 12 – 1997. – p. 267–276.
66. Pozorska J., Scherer M. Company Bankruptcy Prediction with Neural Networks // *Artificial Intelligence and Soft Computing* – 2018. – p.183-189
67. Raikov A. Manufacturer’s Strategic Risk Temperature Assessment with Convergent Approach, Cognitive Modelling and Blockchain Technology // *IFAC-PapersOnLine*. – 2019. – T. 52. – №. 13. – C. 1289-1294.
68. Shi Y., Li X. An overview of bankruptcy prediction models for corporate firms: A systematic literature review // *Intangible Capital* 15(2) – 2019.
69. Shin K.S., Lee T.S., Kim H.J. An application of support vector machines in bankruptcy prediction model // *Expert Systems with Applications* 28 – 2005. – p. 127–135.
70. Soui M., Smiti S., Mkaouer M.W., Ejbali R. Bankruptcy Prediction Using Stacked Auto-Encoders // *Applied Artificial Intelligence* 2019. – p. 80-100.
71. Suss J., Treitel H. Predicting bank distress in the UK with machine learning // *Bank of England Staff Working Paper No. 831* – 2019.
72. Tsai C.F., Wu J.W. Using neural network ensembles for bankruptcy prediction and credit scoring // *Expert Systems with Applications* 34 – 2008. – p. 2639–2649.
73. Tseng F. M., Lin L. A quadratic interval logit model for forecasting bankruptcy // *Omega*. – 2005. – T. 33. – №. 1. – p. 85-91.
74. Wagenmans F. Machine learning in bankruptcy prediction: Master’s thesis. - 2017
75. Wang G., Ma J., Yang S. An improved boosting based on feature selection for corporate bankruptcy prediction // *Expert Systems with Applications* 41 – 2014. – p. 2353–2361.
76. Wang N. Bankruptcy Prediction Using Machine Learning // *Journal of Mathematical Finance* 7 – 2017. – p. 908-918.
77. Zoricak M., Gnip P., Drotar P., Gazda V. Bankruptcy prediction for small- and medium-sized companies using severely imbalanced datasets // *Economic Modelling* 84 – 2020. – p. 165-176.