

УНИВЕРСИТЕТСКИЙ КОНСОРЦИУМ ИССЛЕДОВАТЕЛЕЙ
БОЛЬШИХ ДАННЫХ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ВЯТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

БОЛЬШИЕ ДАННЫЕ И ПРОБЛЕМЫ ОБЩЕСТВА

**Сборник статей по итогам
Международной научной конференции
(Киров, 19–20 мая 2022 г.)**

Ответственный редактор А.Ю. Саркисова

Томск
Издательство Томского государственного университета
2022

УДК 004.6
ББК 16.23
Б79

Большие данные и проблемы общества : сб. статей по итогам
Б79 Международной научной конференции (Киров, 19–20 мая 2022 г.) / отв.
ред. А.Ю. Саркисова. – Томск : Издательство Томского государственного
университета, 2022. – 168 с.

ISBN 978-5-907572-19-5

В сборник включены статьи участников международной научной конференции «Большие данные и проблемы общества» (Киров, 19–20 мая 2022 г.), проходившей в рамках Международного весеннего конгресса Университетского консорциума исследователей больших данных.

Сборник адресован научным сотрудникам, преподавателям и студентам высших учебных заведений, специалистам-практикам в области анализа данных, всем, кто интересуется вопросами использования технологий Big Data для решения научных и прикладных социально значимых проблем.

УДК 004.6
ББК 16.23

ISBN 978-5-907572-19-5

© Авторы статей, 2022

© Томский государственный университет, 2022

СОДЕРЖАНИЕ

<i>Александрова Ю.К., Новожилова А.Е., Петров Е.Ю.</i> ОЦЕНКА РАСПРОСТРАНЁННОСТИ АКАДЕМИЧЕСКОГО МОШЕННИЧЕСТВА НА ОСНОВЕ КЛАСТЕРНОГО АНАЛИЗА ТЕКСТОВ СООБЩЕНИЙ В УНИВЕРСИТЕТСКИХ ПАБЛИКАХ СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ»	5
<i>Барило И.И.</i> ПОИСК НОВЫХ МАЙНИНГ-ПУЛОВ ДЛЯ ОПЕРАТИВНОГО ОТСЛЕЖИВАНИЯ ПОЯВЛЕНИЯ НОВЫХ КРИПТОВАЛЮТ	11
<i>Барышев А.А., Сербина Г.Н.</i> ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ СИСТЕМАТИЧЕСКОГО АНАЛИЗА ИССЛЕДОВАНИЙ СОЦИАЛЬНЫХ ФЕНОМЕНОВ: ПРОБЛЕМА ОБЕСПЕЧЕНИЯ ГОМОГЕННОСТИ ЗНАЧЕНИЙ КОНЦЕПТОВ	20
<i>Беляева Е.С., Петрова И.Э.</i> СЕЛЬСКАЯ ШКОЛА: АНАЛИЗ ПРЕДСТАВЛЕНИЯ ПРОБЛЕМ НА ПУБЛИЧНЫХ АРЕНАХ	29
<i>Благинин А.Л., Сайфулин Э.Р., Саркисова А.Ю.</i> ИЗ ОПЫТА ОРГАНИЗАЦИИ АВТОМАТИЗИРОВАННОГО СБОРА ДАННЫХ В ТОМСКОМ УНИВЕРСИТЕТЕ	34
<i>Бушкова-Шиклина Э.В., Москвин А.С., Кушова И.А., Голенок М.П., Одегова К.И.</i> ПРОБЛЕМА ИЗМЕРЕНИЯ ВИРТУАЛЬНОЙ КУЛЬТУРНОЙ АКТИВНОСТИ В СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ» МЕТОДАМИ BIG DATA	47
<i>Галкин С.А., Найдено Д.Г., Гойко В.Л.</i> ИССЛЕДОВАНИЕ УРОВНЯ ТРЕВОЖНОСТИ, СВЯЗАННОГО С COVID-19: ВЛИЯНИЕ СОЦИАЛЬНО-ДЕМОГРАФИЧЕСКИХ И ЛИЧНОСТНЫХ ХАРАКТЕРИСТИК	56
<i>Гмызина Э.В., Городилова Т.С.</i> МЕТААНАЛИЗ ПОНЯТИЯ «КУЛЬТУРНЫЙ КОД»	63
<i>Иванова Т.Н., Новичкова А.С.</i> РОСДИСТАНТ КАК ИННОВАЦИОННЫЙ КОНЦЕПТ НОВОГО ПОКОЛЕНИЯ ОБУЧЕНИЯ СТУДЕНТОВ ВУЗОВ В ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ	69
<i>Исакова И.А.</i> ВОЗМОЖНОСТИ ОЦЕНКИ КРЕАТИВНОСТИ ПО ЦИФРОВЫМ СЛЕДАМ	75
<i>Кашир В.В., Барышев А.А., Губанов А.В., Мищенко Е.С.</i> ПРАГМАТИКА ДЕСИГНАТОВ БЛАГОТВОРИТЕЛЬНОСТИ: ОПЫТ СЕТЕВОГО АНАЛИЗА АНГЛОЯЗЫЧНЫХ НАУЧНЫХ ТЕКСТОВ	80
<i>Корзун Д.Ж.</i> О ЦИФРОВОМ АССИСТЕНТЕ ДЛЯ МОНИТОРИНГА ЖИЗНЕСТОЙКОСТИ ЧЕЛОВЕКА В УСЛОВИЯХ ПОВСЕДНЕВНОЙ ЖИЗНИ	86
<i>Косьмина Е.А.</i> БОЛЬШИЕ ДАННЫЕ В КОМПЬЮТЕРНОМ СПОРТЕ	90

<i>Крюков В.В., Завалин Г.С., Гмарь Д.В., Батурина О.А., Цуранов Э.В.</i> ПРИМЕНЕНИЕ УЧЕБНОЙ АНАЛИТИКИ ДЛЯ УПРАВЛЕНИЯ «УСПЕШНОСТЬЮ» СТУДЕНТА	94
<i>Кузнецова Н.М.</i> ФОРМИРОВАНИЕ КОМПЕТЕНЦИЙ БУДУЩЕГО В УСЛОВИЯХ ОРГАНИЗАЦИОННОЙ НЕОПРЕДЕЛЕННОСТИ	100
<i>Лагутин Ю.В.</i> СЦИЕНТИЗАЦИЯ БЫТА В РОССИЙСКИХ МЕГАПОЛИСАХ: ЭТАПЫ ВНЕДРЕНИЯ ИННОВАЦИЙ С 2000 ПО 2022 г.	106
<i>Мищенко Е.С., Барышев А.А., Губанов А.В.</i> АНАЛИЗ ДИСКУРСОВ БЛАГОТВОРИТЕЛЬНОСТИ ЧЕРЕЗ ПРИЗМУ БИБЛИОМЕТРИЧЕСКОЙ СЕТИ	111
<i>Мосунова Л.А., Митягина Е.В.</i> ПРОБЛЕМА ИЗУЧЕНИЯ КУЛЬТУРНОГО КОДА РОССИЯНИНА С ПОМОЩЬЮ БОЛЬШИХ ДАННЫХ	121
<i>Петрищева М.А., Александрова Ю.К.</i> МНЕНИЯ О «ЕДЕ БУДУЩЕГО» В ЗЕРКАЛЕ АНАЛИЗА ДАННЫХ СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ»	126
<i>Ситникова И.В.</i> ПРОФЕССИОНАЛЬНО-ОБРАЗОВАТЕЛЬНЫЕ СТРАТЕГИИ СТУДЕНЧЕСКОЙ МОЛОДЕЖИ	135
<i>Стефановский Д.В., Басина П.А.</i> ИДЕНТИФИКАЦИЯ ЯЗЫКА ПОЛЬЗОВАТЕЛЕЙ «ВКОНТАКТЕ»: РАЗВЕДОЧНЫЙ АНАЛИЗ	139
<i>Тимирьянова В.М., Лакман И.А., Баймурзина Г.Р.</i> ИССЛЕДОВАНИЯ НА «БОЛЬШИХ ДАННЫХ»: ПРИМЕНЯТЬ ЛИ АГРЕГАЦИЮ В ЦЕЛЯХ МОДЕЛИРОВАНИЯ?	144
<i>Фабрикант М.С.</i> БОЛЬШИЕ ДАННЫЕ В КРОСС-КУЛЬТУРНЫХ ИССЛЕДОВАНИЯХ: ДОВЕРИЕ К ИНФОРМАЦИИ И ДОВЕРИЕ К РЕЗУЛЬТАТАМ	150
<i>Ярмак О.В., Ткаченко А.Ю.</i> ПРОФЕССИОНАЛЬНЫЕ ОРИЕНТИРЫ СЕЛЬСКОЙ МОЛОДЁЖИ КАК ФАКТОР МИГРАЦИИ В ГОРОД	157
СВЕДЕНИЯ ОБ АВТОРАХ	161
ABOUT THE AUTHORS	165

ОЦЕНКА РАСПРОСТРАНЁННОСТИ АКАДЕМИЧЕСКОГО МОШЕННИЧЕСТВА НА ОСНОВЕ КЛАСТЕРНОГО АНАЛИЗА ТЕКСТОВ СООБЩЕНИЙ В УНИВЕРСИТЕТСКИХ ПАБЛИКАХ СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ»

Ю.К. Александрова, А.Е. Новожилова, Е.Ю. Петров

Национальный исследовательский Томский государственный университет, Томск, Россия

Аннотация. В настоящем сообщении методами анализа больших данных, а именно кластеризации текстов, анализируется распространённость практик академического мошенничества в российских вузах на основе содержания постов и комментариев в социальной сети «ВКонтакте». Авторы оценивают долю обсуждений, связанных с академической нечестностью, динамику количества соответствующих сообщений в течение 2020–2021 учебного года и их содержание. Выявлено, что сегмент обсуждения вопросов, связанных с академическим мошенничеством, составляет более 1/10 всех переписок в университетских пубliках социальной сети «ВКонтакте» и примерно в равном количестве представлен предложением (рекламой) соответствующих услуг от исполнителей и спросом на них со стороны студентов. Отмечена более высокая распространённость данной темы в ведущих вузах, включая национальные исследовательские и федеральные университеты.

Ключевые слова: социальные сети, ВКонтакте, академическое мошенничество, академическая нечестность, большие данные, кластерный анализ

В связи с пандемией COVID-19 в 2020 г. организация учебного процесса в российских вузах претерпела резкие изменения на разных уровнях. Переход на дистанционные формы обучения, по оценкам преподавателей, в частности, привёл к росту академической нечестности: студенты стали чаще прибегать к обману, который в условиях дистанционного контроля знаний сложнее отслеживать [1].

Специфика изучения темы академической нечестности заключается в том, что студенты, заказывающие работы у сторонних лиц и организаций, предпочитают сохранять анонимность, а работа исполнителей во многих случаях неформальна, выполняется без регистрации и уплаты налогов. Несмотря на обширное количество исследований по данной тематике [2–4], все они проведены социологическими методами. Однако использование опросов и экспериментов при изучении академического мошенничества имеет ряд значимых ограничений: боязнь студентами социального осуждения может привести к намеренному искажению ответов, исследования охватывают сравнительно небольшие, локальные выборки (отдельные вузы, регионы), проблема академического мошенничества рассматривается только с точки зрения спроса на соответствующие услуги без учёта сегмента их предложения.

В связи с этим в рамках настоящего исследования в качестве источника данных о распространённости спроса и предложения услуг по выполнению различных учебных работ для студентов выбраны тексты сообщений в университетских пубliках/группах в социальной сети «ВКонтакте», где данная тема обсуждается активно и открыто. Кроме того, в связи с вынужденной самоизоляцией многие процессы перешли в онлайн-среду, в том числе и рассматриваемое явление образовательного процесса. С помощью методов анализа больших данных авторы попытались получить неискажённую информацию относительно объема академического мошенничества и его содержания.

В настоящей работе в качестве проявлений академического мошенничества будут рассматриваться спрос и предложение от исполнителей по выполнению различных видов работ: написание курсовых / дипломных работ, решение задач, выполнение лабораторных работ, замена студентов на экзамене / лекции, заполнение тетрадей по практике, продажа ответов / решений и др., то есть всё, что связано с фальсификацией результатов работы обучающихся.

Таким образом, цель настоящей работы – отразить объективное состояние распространённости академического мошенничества в современной российской университетской среде (в зеркале его репрезентации в социальной сети).

Актуальность исследования определяется: 1) распространённостью академического мошенничества как одной из ключевых проблем современного высшего образования; 2) значением социальных сетей в студенческой коммуникации, формировании общественного мнения, репрезентации интересов, точек зрения социальных групп; 3) введением в научный оборот большого эмпирического материала, полученного с использованием современных технологий сбора и обработки данных.

Материалы и методы

Для настоящего исследования выбраны 1 746 университетских пабликов/групп в социальной сети «ВКонтакте», объединяющих студентов и выпускников российских вузов: «Подслушано МГТУ им. Баумана», «МГУ имени М.В. Ломоносова», «Типичный МАИ», «Говорит НИУ ВШЭ СПб», «Молодёжка ТюмГУ», «Студенческий клуб ВятГУ / Студклуб», «MGIMO OVERHEARD», «Признавайки Политех (ИрНИТУ)» и др. Из них через портал по работе с данными социальных сетей, разработанный Национальным исследовательским Томским государственным университетом [5] проведена выгрузка постов (сообщений на стене группы) и комментариев за 2020/21 учебный год. Общее количество исследованных текстов составило более 740 тыс. постов и комментариев (в «рабочую» выгрузку не включались сообщения, которые содержали только визуальный контент, а также текстовые сообщения, содержащие менее 15 символов). В анализ включены паблики 696 вузов и филиалов из всех регионов Российской Федерации, а также отдельные филиалы в Беларуси, Казахстане, Узбекистане.

Содержательный анализ такого объема текстов возможен с применением инструментов анализа больших данных. В настоящем исследовании использована программная платформа для интеллектуального текстового анализа и анализа данных PolyAnalyst (разработчик – Megarputer Intelligence) [6]. Для группировки текстов по их тематике реализована цепочка последовательных кластеризаций методом вероятностного латентно-сематического анализа.

Тексты постов и комментариев в социальных сетях неоднородны по содержанию, стилю и размеру. Исследуемый массив данных содержит как сообщения с чётко определяемой тематикой – например, *«Скажите пожалуйста, кто может помочь с курсовыми по менеджменту, одну точно до понедельника ?? Анон»* (академическое мошенничество), *«Внимание, вожатые, методисты, старшие вожатые, стажеры, и все-все причастные к романтике вожатского лета! Продолжается прием заявок на Всероссийский конкурс «Лига вожатых», который проводится в целях реализации и выполнения федерального проекта «Учитель будущего» национального проекта «Образование»* (студенческие мероприятия и организации); так и смешанной либо сложно определяемой тематикой – *«Город засыпает... Просыпаются заочники...»*, *«Пятница, это кайф. Анон.»*, *«Вечером кто тусить?»* (большинство подобных сообщений отнесены к кластеру «Общение, знакомства, юмор»).

После каждого этапа кластеризации на основании экспертной оценки отделялись «чистые» кластеры, сформированные из текстов с корректно определённой тематикой, и «смешанные» кластеры, к которым алгоритмы машинного обучения отнесли тексты различного содержания. Вторая группа текстов кластеризовалась повторно. Такой подход позволил повысить точность отнесения текста к определенной тематике и выявить более мелкие подтемы, значимые для исследования.

Кластеры, сформированные программой PolyAnalyst из ключевых слов и лексем, переименованы вручную, близкие темы объединены. Методом таксономии (поиска по ключевым словам) сообщения, не относящиеся к исследуемой теме, дополнительно отфильтрованы и исключены из анализируемого массива.

Степень распространённости академического мошенничества

Вопросы спроса и предложения, касающиеся академического мошенничества, вошли в топ-5 наиболее обсуждаемых тем в университетских группах / пабликах в социальной сети «ВКонтакте», уступив по объему сообщений только собственно учебным вопросам, общению, теме поступления и перевода (рис. 1).

Менее значимыми по числу сообщений оказались также общежитие и бытовые вопросы, экзамены и сессия, стипендии, справки и правовые вопросы, студенческие мероприятия и организации. Сегмент обсуждений анализируемой тематики составил 11,4 % от общей массы сообщений.



Рис. 1. Распределение текстов постов и комментариев по тематикам, шт.

Динамика количества сообщений по тематике академического мошенничества повторяет динамику всех опубликованных в университетских пабликах сообщений с более заметными пиками активности в сессионные периоды – в эти периоды обсуждение исследуемой тематики у студентов выходит на первый план (рис. 2).

В рамках исследования также проведен сравнительный анализ доли сообщений, отнесённых к кластеру «Академическое мошенничество», по группам вузов: национальные исследовательские университеты (29), опорные вузы (15), федеральные вузы (10) и прочие вузы, не отнесённые к другим категориям (642) (рис. 3).

Наименьшая доля сообщений, посвящённых покупке и продаже студенческих работ, отмечена среди «прочих» вузов (10,2%).

Среди федеральных и национальных исследовательских университетов доля сообщений по тематике академической нечестности в общем количестве сообщений значительно превысила показатель по прочим вузам, не входящим ни в одну из категорий, и составила 13,9 и 18,6% соответственно.

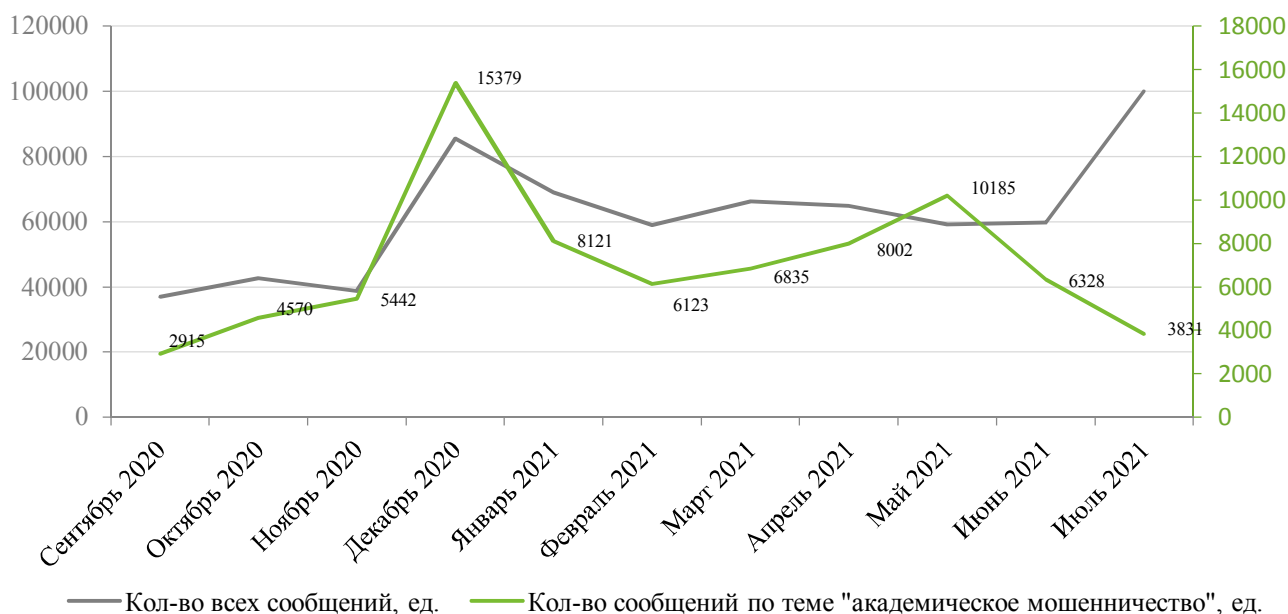


Рис. 2. Динамика количества сообщений

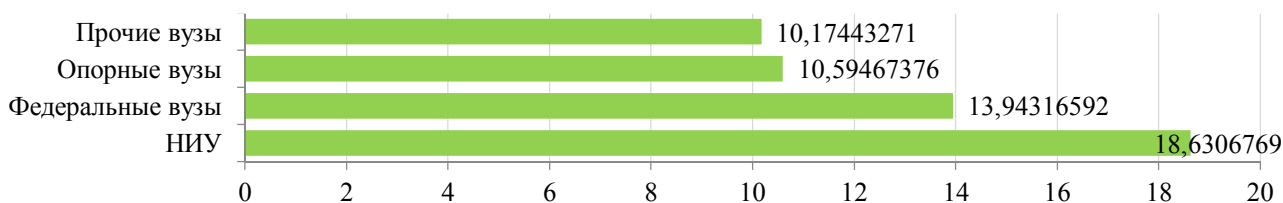


Рис. 3. Доля сообщений кластера «Академическое мошенничество» в общем количестве сообщений по группам вузов, %

Полученные результаты свидетельствуют, что проблема академического мошенничества характерна не только для вузов с невысоким рейтингом и низким качеством обучения, но также и для ведущих вузов страны. Примечательно, что для наиболее рейтинговых вузов со статусом национального исследовательского университета значимость темы академического мошенничества в общем объёме сообщений оказалась максимальной.

Виды академического мошенничества

Более детальный анализ сообщений, выделенных в кластер «Академическое мошенничество», позволил выявить 8 подтем (рис. 4).



Рис. 4. Доля сообщений (постов и комментариев) по подтемам в общем количестве сообщений, отнесённых к тематике академического мошенничества, %

Наибольшее количество сообщений являются предложениями от исполнителей и рекламой соответствующих услуг – они преобладают в комментариях (при просьбе о помощи исполнители пишут свои предложения в комментариях к посту). Также выделены следующие подтемы:

- запросы на помощь по отдельным предметам («Срочно ищу человека кто шарит в компьютерной графике и знает что требует рещетов. Анон»);
- решение задач («Есть рещалы по химии? нужно рещить задания по органике»);
- написание рефератов, курсовых и дипломных работ («Нужна помощь в написании курсового проекта по предмету «моделирование социально-экономических систем». Анон»);
- купля-продажа тетрадей, методичек и готовых ответов («Куплю готовую тетрадь по лрд с рисунками»);
- обсуждение исполнителей (описание своего опыта взаимодействия с исполнителями, жалобы и предупреждения пользователей соцсетей о мошенничестве в этой сфере услуг) («Осторожно с предоплатами неизвестно кому! Работают фейки-боты, мошенники на доверии!»);
- помощь с прохождением практики («Кто сделает отчёт по практике?»);
- замена на лекциях и помощь с экзаменом («Здравствуйте. Кто может помочь с экзаменом по инженерной графике у Ширшовой?», «Нужна девочка 5-6 курс для сдачи нб по терапии. Анонимно», «Схожу на лекции в любое время. Писать в лс.»).

Из результатов исследования следует, что проблема академического мошенничества масштабна и во многом недооценена. Вопросы покупки и продажи студенческих работ входят в топ-5 наиболее обсуждаемых тем в университетских группах / пабликах в социальной сети «ВКонтакте». Сегмент обсуждения вопросов, связанных с академическим мошенничеством, составляет более 1/10 всех переписок студентов в университетских сообществах, при том, что существенная доля сообщений не связана с учёбой: в переписках обсуждаются общепития, студенческие мероприятия, фильмы, музыка, праздники, публикуются рекламные объявления и т.д.

Интересный вывод состоит в том, что проблема академического мошенничества характерна для всех вузов, включая ведущие вузы страны. При этом максимальной доля сообщений по теме академической нечестности в общем массиве переписок оказалась среди высокорейтинговых национальных исследовательских университетов.

Содержание запросов, публикуемых в постах и комментариях университетских пабликов, свидетельствует о разнообразии видов академической нечестности, к которым прибегают студенты в процессе обучения: решение задач, работа в специальных программах, написание рефератов, дипломных, курсовых работ, выполнение расчетно-графических заданий, подготовка чертежей, прохождение тестов, написание лекций, замена на лекциях и экзаменах, подготовка отчётов по практике, продажа тетрадей и ответов.

Основным ограничением настоящего исследования является обезличенность данных, что не позволяет изучать мотивацию студентов и соотносить факты академической нечестности с учебными успехами и личностными характеристиками обучающихся.

В развитие настоящего исследования можно (а) рассмотреть проблему академической нечестности в динамике для выявления степени влияния на распространённость перехода данного явления на дистанционные формы обучения в связи с пандемией COVID-19; (б) изучить соотношение спроса и предложения соответствующих услуг, а также структуру спроса в разрезе учебных дисциплин и направлений обучения.

Список источников

1. Списать или нет – вот в чем вопрос. URL: <http://libera.media/spisyvanieinterview> (дата обращения: 08.05.2022).
2. Watson G., Sottile J. Cheating in the Digital Age: Do Students Cheat More in Online Courses? // Online Journal of Distance Learning Administration. 2010. № 13 (1). URL: <https://www.westga.edu/~distance/ojdla/spring131/watson131.html> (дата обращения: 08.05.2022).

3. Muafia M., Ahmad Z., Shahzadi E. A Study of Academic Dishonesty of University Students // Proc. 8th International Conference on Recent Advances in Statistics. Lahore. Pakistan. Feb. 8–9, 2011. P. 285–294. URL: <https://www.researchgate.net/publication/266736636> (дата обращения: 08.05.2022).
4. Малошонок Н.Г., Семенова Т.В., Терентьев Е.А. Учебная мотивация студентов российских вузов: возможности теоретического осмысления // Вопросы образования. 2015. № 3. С. 92–121.
5. Портал по работе с данными социальных сетей. URL: <https://lk.opendata.university/> (дата обращения: 05.04.2022).
6. Программная платформа для интеллектуального текстового анализа и анализа данных PolyAnalyst. URL: <https://www.megaputer.com/ru/polyanalyst/> (дата обращения: 20.04.2022).

ASSESSMENT OF THE PREVALENCE OF ACADEMIC FRAUD BASED ON CLUSTER ANALYSIS OF MESSAGES IN UNIVERSITY PUBLICS OF THE SOCIAL NETWORK VKONTAKTE

Yu.K. Alexandrova, A.E. Novozhilova, E.Yu. Petrov

National Research Tomsk State University, Tomsk, Russian Federation

Abstract. The study estimates the prevalence of academic fraud practices in Russian universities on the data of posts' and comments' texts, published in the VKontakte social network, using methods of big data analysis, namely text cluster analysis. The authors evaluate the share of discussions related to academic dishonesty, the dynamics of the number of relevant messages during the 2020–2021 academic year and their content. They reveal that the segment of discussion of issues related to academic fraud accounts for more than 1/10 of all messages in the university public pages of the VKontakte social network and is approximately equally represented by the offer (advertising) of relevant services and the demand for them from students. A higher prevalence of this topic is revealed in leading universities, including national research and federal universities.

Keywords: social networks, VKontakte, academic fraud, academic dishonesty, big data, cluster analysis

ПОИСК НОВЫХ МАЙНИНГ-ПУЛОВ ДЛЯ ОПЕРАТИВНОГО ОТСЛЕЖИВАНИЯ ПОЯВЛЕНИЯ НОВЫХ КРИПТОВАЛЮТ

И.И. Барило

Костромской государственной университет, Кострома, Россия

В статье изложена актуальная проблема поиска новых криптовалют, рассмотрены основные понятия, относящиеся к их добыче, существующие способы получения информации, популярные информационные сервисы, публикующие информацию о криптовалютах, а также пути распространения информации о криптовалюте от ее разработчиков. Предложен подход к поиску новых майнинг-пулов, рассмотрены этапы их автоматического нахождения и особенности проверки корректности их определения, показаны возможности получения информации с майнинг-пулов, а также использования их для поиска новых криптовалют. Проведен поиск всех пулов предложенным способом и подготовлен их список с указанием частоты их использования при добыче различных криптовалют. Приведен практический пример использования предложенного подхода, когда новая криптовалюта была найдена ранее одного из популярных сервисов.

Ключевые слова: криптовалюта, пул, майнинг, регулярные выражения, поиск новых криптовалют

В настоящее время активно развивается криптовалютный рынок, постоянно появляются новые криптовалюты, которые предлагают различный функционал и возможности их интегрирования в различные сферы деятельности.

Одними из них являются криптовалюты, которые добываются (проводятся математические расчеты) с помощью Proof of work, то есть доказательств работы, либо других вариантов их добычи, которые частично используют Proof of work. Существует взаимосвязь их известности и сложности их добычи.

При появлении новой криптовалюты она мало известна, сложность ее добычи низкая, но по мере ее развития, добавления ее на биржи и внедрения в проекты о ней узнает все больше людей, добывающих криптовалюты, которые начинают присоединяться к ее добыче, что значительно повышает сложность добычи монет этой криптовалюты.

В качестве примера на рис. 1 [1] приведена сложность сети одной из наиболее известных криптовалют – Bitcoin.

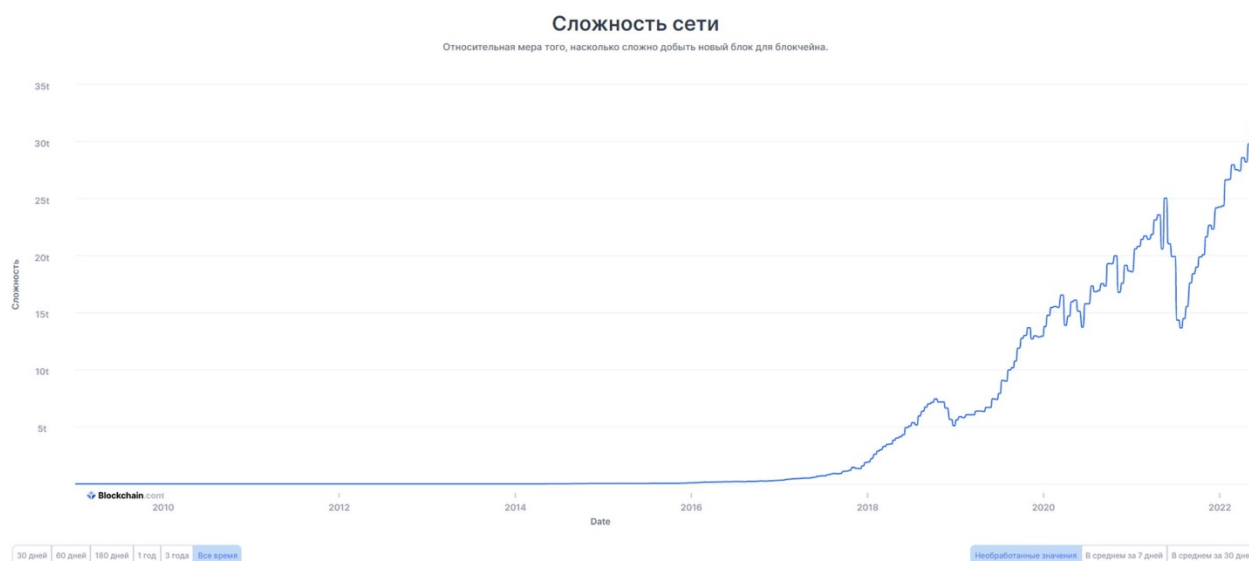


Рис. 1. Сложность сети

По графику видно, что в самом начале сложность сети была очень мала, что позволяло добывать ее в большом количестве на относительно слабом компьютерном оборудовании, даже, например, на ноутбуках. Сейчас, из-за крайне высокой сложности ее добычи (расчетов), данную криптовалюту добывают исключительно на специализированном оборудовании: ноутбуки и компьютеры крайне неэффективны в ее добыче. Что позволяет сделать вывод – чем раньше мы узнаем о появлении новой криптовалюты, тем проще ее будет добыть.

Следует отметить, что много новых криптовалют быстро теряют свою ценность после выхода на криптовалютные биржи, также они могут просто закрыться разработчиками, или могут возникнуть другие технические проблемы, поэтому добыча новых криптовалют – это большой риск.

Одним из факторов добычи криптовалюты являются майнинг-пулы.

«Майнинг-пул – это группа майнеров, которые делятся своими вычислительными мощностями по сети и получают вознаграждение в зависимости от количества мощности, которое каждый вносит, а не от того, найдет ли пул блок. Пулы для майнинга помогают сделать доход майнеров более предсказуемым» [2].

На рис. 2 представлен график распределения вычислительной мощности, также называемой хэшрейт, среди пулов для криптовалюты Биткоин. На рисунке видно, что большую часть вычислений выполняют именно пулы.



Рис. 2. Оценка распределения хэшрейта среди крупнейших майнинговых пулов

Существует два основных варианта поиска новых криптовалют:

1. Информация от разработчиков криптовалют (разработчики публикуют информацию о своих криптовалютах, особенностях и возможности их добычи).
2. Информационные сервисы.

Первый вариант поиска криптовалют – у разработчиков криптовалют, которые публикуют информацию о своих криптовалютах, особенностях и возможностях ее добычи на специально созданном ими сайтах, на которых изложена полная информация о них, но найти такие сайты через поисковые системы на раннем этапе их создания затруднительно.

В качестве примера взята криптовалюта – Ravencoin, которая на момент подготовки статьи уже существует значительное время.

Информацию по данной криптовалюте можно посмотреть на сайте ravencoin.org [3] (рис. 3).

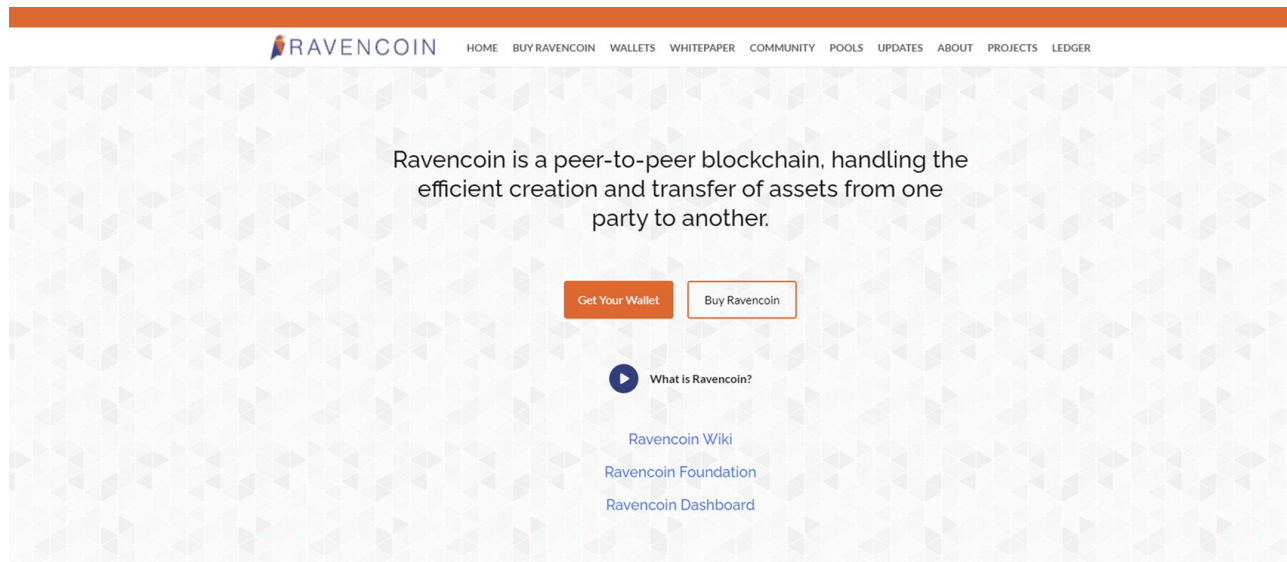


Рис. 3. Главная страница сайта криптовалюты Ravencoin

Помимо сайта, разработчики, как правило, публикуют информацию на популярном форуме Bitkointalk. На форуме есть ряд разделов, наиболее популярным является англоязычный раздел форума [4], посвященный новым криптовалютам.

На рис. 4 представлен фрагмент описания криптовалюты Ravencoin из сообщения на форуме Bitkointalk [5].



Рис. 4. Фрагмент описания криптовалюты Ravencoin из сообщения на форуме Bitkointalk

Следует обратить внимание, что заблокирован доступ к части форума Bitkointalk, согласно статье в журнале bits.media [6], поэтому интернет-провайдерами может быть ограничен доступ к ряду других страниц форума, доступ к которым не должен был быть ограничен, включая страницу с описанием криптовалюты Ravencoin.

Вторым путем поиска новых криптовалют являются специализированные сервисы. Существуют крупные информационные сервисы, например, coinmarketcap.com [7], но на нем как правило появляется информация об уже зарекомендовавших себя криптовалютах, и новые, только появившиеся, не отображаются в нем.

На рис. 5 представлена страница [8] с информацией о Ravencoin.

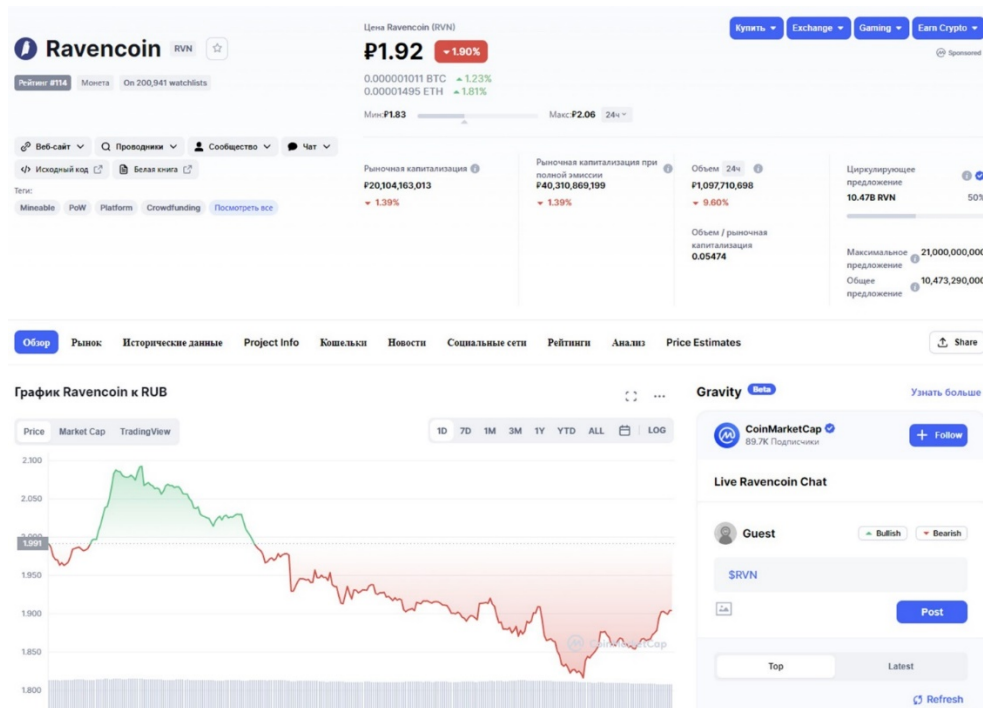


Рис. 5. Информация о Ravencoin на сайте coinmarketcap.com

Также есть сервисы, которые отслеживают информацию с майнинг-пулов и могут сообщать о появлении на них новых криптовалют. Примером такого сервиса является сайт miningpoolstats.stream, на котором есть отдельная страница [9] с новыми криптовалютами (рис. 6).

New PoW Coins		Algorithm	Estimated Age	Current Height	Pools (known)	Pools Hashrate	Network Hashrate	Network Difficulty
1.	Reflect	Script	4 days	7121	5	4.24 Gh/s	110.82 Gh/s	63.48512
2.	Safeminex	YescryptR16	6 days	9112	9	860.54 Mh/s	862.20 Mh/s	0.01458139
3.	TubePay	Script	6 days	5034	7	1.28 Gh/s	1.89 Gh/s	195.374
4.	Verbas	GhostRider	7 days	5331	12	2.83 Mh/s	2.68 Mh/s	0.07779051
5.	Picacoin	SHA-256	29 days	43038	1	0.00 Mh/s	63.44 Gh/s	842.04 K
6.	OjaCoin	Xevan	33 days	47920	4	672.26 Mh/s	753.02 Mh/s	11.37518
7.	Shahpepy	X16Rv2	40 days	58498	9	315.46 Mh/s	319.95 Mh/s	4.50224173
8.	Itasecoin	Script	41 days	59790	12	2.57 Gh/s	2.57 Gh/s	35.02943
9.	Argoad	Script	41 days	119925	7	3.83 Gh/s	4.53 Gh/s	31.40059
10.	Bitoreum	GhostRider	41 days	30044	18	8.88 Mh/s	9.25 Mh/s	0.25583819
11.	Bitcoin Zero	Lyra2z	42 days	24584	5	15.58 Mh/s	8.12 Mh/s	0.37254846
12.	CoffeeCoin	Script	44 days	63362	0	0.00 Mh/s	0.00 Mh/s	0.00000000
13.	Reacoin	YescryptR16	55 days	158786	21	379.14 Mh/s	397.08 Mh/s	0.00230540
14.	89Coin	Script	55 days	15899	1	0.00 Mh/s	149.21 Mh/s	1.05 K
15.	Canada vCoin	Script	59 days	17071	2	2.16 Th/s	71.25 Th/s	4.82 M

Рис. 6. Страница с новыми криптовалютами на сайте miningpoolstats.stream

У таких сервисов следует учитывать ряд недостатков:

- возможна задержка публикации найденных криптовалют для пользователей с бесплатным доступом к сервису;
- могут не отслеживаться новые пулы, на которых появилась новая криптовалюта.

Следует учесть, что разработчики криптовалют, как правило, сами создают пул для ее добычи, и после публикации на форуме данную криптовалюту могут добавить популярные пулы. Зачастую популярные пулы не сразу подключают добычу через них новых криптовалют, чему есть ряд причин, например, специалисты, управляющие пулом, не успевают их отследить, не уверены в надежности криптовалюты, опасаются за безопасность программного обеспечения данной криптовалюты для установки на сервер пула.

Также существуют форумы, где можно найти информацию, публикуемую разработчиками криптовалют.

Самым известным из таких форумов является Bitkointalk, помимо описания самой криптовалюты, разработчики указывают как ее добывать, обычно приводя примеры командной строки для программ добычи криптовалют, в которой также указан адрес сайта-пула для ее добычи. Примеры настроек для программы добычи криптовалюты приведены на рис. 7.

```
Mining Pools  
MinerMore: https://minermore.com/  
code: stratum+tcp://pool.minermore.com:4501 -u WALLEТ_ADDRESS -p x ≠ USA  
code: stratum+tcp://eu.minermore.com:4501 -u WALLEТ_ADDRESS -p x ≠ Europe  
code: stratum+tcp://hk.minermore.com:4501 -u WALLEТ_ADDRESS -p x ≠ Asia  
  
Minepool: https://www.minepool.com/  
code: stratum+tcp://stratum.minepool.com:3636 -u WALLEТ_ADDRESS -p c=RVN  
  
YiiMP.eu: http://yiiimp.eu/  
  
CryptoPool.PARTY: https://cryptopool.party/  
code: stratum+tcp://cryptopool.party:3636 -u <WALLEТ_ADDRESS> -p c=RVN  
  
IceMining.ca: https://icemining.ca/site/gocoins?id=2089  
code: -a x16r -o stratum+tcp://mine.icemining.ca:3636 -u <WALLEТ ADDRESS> -p c=RVN  
  
Coinblockers: https://rvn.coinblockers.com/  
code: US Stratum: stratum+tcp://rvn-us.coinblockers.com:4449 -u  
WALLEТ_ADDRESS.yourRigName For EU stratum change "us" to "eu"  
  
gos.cx: https://gos.cx/site/GetCoin?c=RVN  
code: -a x16r -o stratum+tcp://eu.gos.cx:3637 -u wallet.rigName -p c=RVN - Europe  
code: -a x16r -o stratum+tcp://us.gos.cx:3637 -u wallet.rigName -p c=RVN - USA  
code: -a x16r -o stratum+tcp://ru.gos.cx:3637 -u wallet.rigName -p c=RVN - RU
```

Рис. 7. Фрагмент описания криптовалюты Ravencoin из сообщения на форуме Bitkointalk

Пул может предоставить информацию о криптовалютах, которые на нем добываются. В качестве примера приведем gos.cx [10] из указанного на форуме описания криптовалюты RVN. Главная страница пула с информацией о добываемой на нем криптовалюте приведена на рис. 8.

У данного пула есть возможность получения информации о криптовалютах, добываемых на нем через специальные API-запросы, возвращающие json-файл содержащих требуемую информацию [11]. На странице пула со служебной информацией, представленной на рис. 9, приведен ряд примеров с запросами к пулу для получения такой информации.

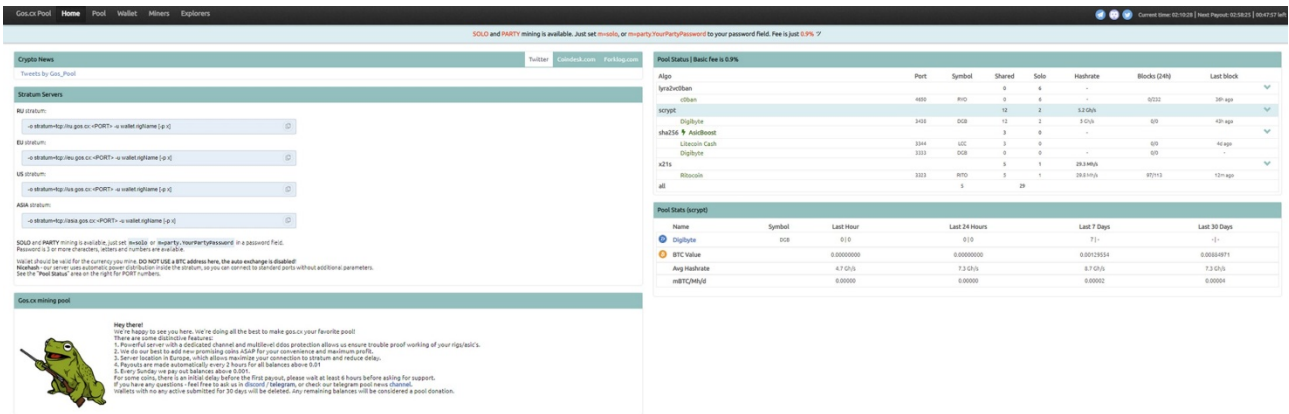


Рис. 8. Главная страница пула gos.cx

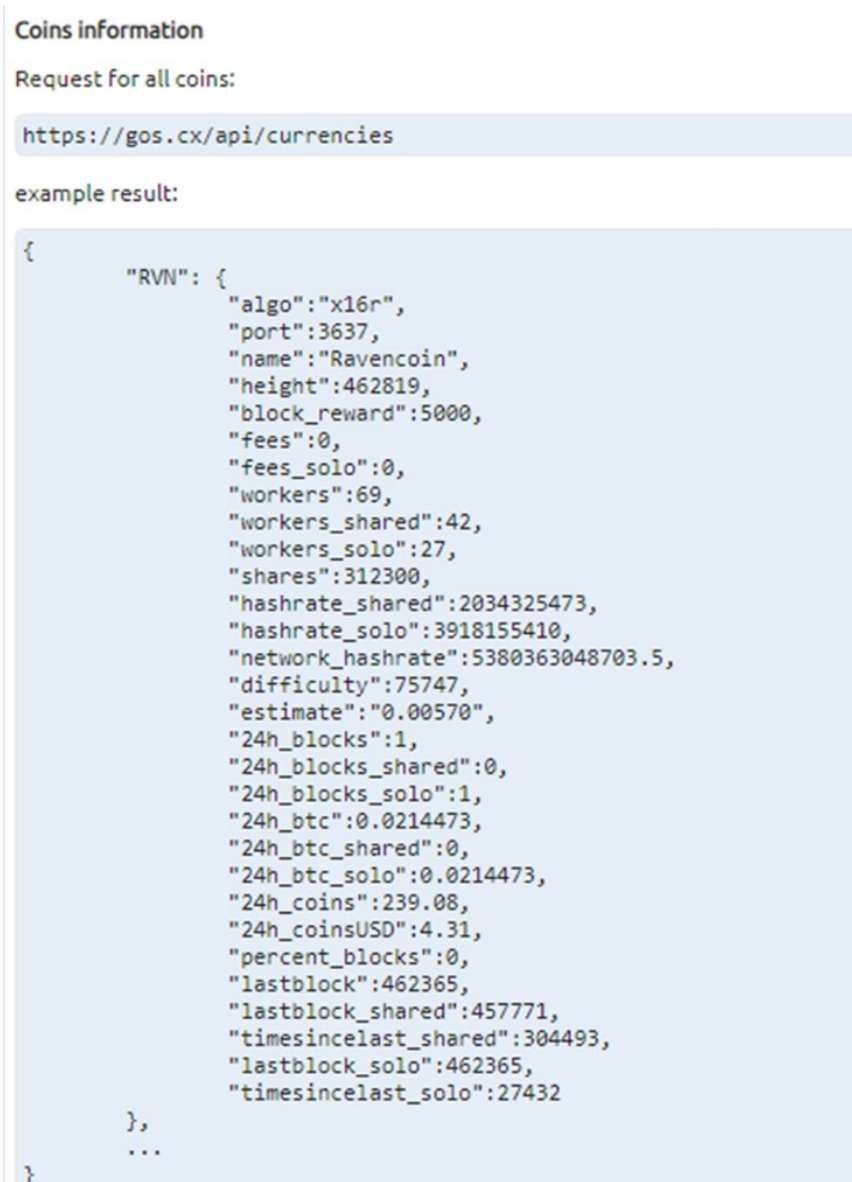


Рис. 9. Фрагмент страницы с примерами API-запросов

Большое количество майнинг-пулов построено на исходном коде проекта uimr [12], следовательно, и API-запросы к ним одинаковые, за редким исключением, если были внесены существенные изменения в функционал пула.

Информация, размещенная разработчиками на форуме, позволяет проводить автоматический поиск названий пулов в описании криптовалют. На рис. 10 видно, что можно определить название сайта пула через регулярное выражение, отслеживающее приставку “stratum+tcp”, после которой идет название сайта. Следует учесть, что порой для сайта используются субдомены, формата “eu”, “us” и пр. – которые следует убрать, либо протестировать получение информации с таким субдоменом и без него.

```
gos.cx: https://gos.cx/site/GetCoin?c=RVN
code: -a x16r -o stratum+tcp://eu.gos.cx:3637 -u wallet.rigName -p c=RVN - Europe
code: -a x16r -o stratum+tcp://us.gos.cx:3637 -u wallet.rigName -p c=RVN - USA
code: -a x16r -o stratum+tcp://ru.gos.cx:3637 -u wallet.rigName -p c=RVN - RU
```

Рис. 10. Фрагмент описания криптовалюты Ravencoin из сообщения на форуме Bitkointalk

Получив название сайта, к нему следует добавить “/api/currencies” и проверить на получение ответа в виде json-файла. Следует учесть, что протокол может быть как https, так и http.

В случае, например, найденного фрагмента “stratum+tcp://eu.gos.cx” для него будет извлечен адрес сайта “eu.gos.cx”, на основе которого будет сформирован следующий список вариантов обращений к сайту:

- 1) “https://eu.gos.cx/api/currencies”;
- 2) “http://eu.gos.cx/api/currencies”;
- 3) “https://gos.cx/api/currencies”;
- 4) “http://gos.cx/api/currencies”.

После автоматического опроса всех вариантов будет определен верный – “https://gos.cx/api/currencies”, который и будет использован в дальнейшем для периодических опросов для поиска появившихся на нем новых криптовалют.

Рассмотрим json-файл ответа, полученного на запрос “https://gos.cx/api/currencies”.

Фрагмент json-файла, полученного от пула:

```
{
  "DGB": {
    "algo": "scrypt",
    "port": 3438,
    "name": "Digibyte",
    "height": 15076861,
    "block_reward": "454.49002805",
    "fees": 0.9,
    "fees_solo": 0.9,
    "workers": 15,
    "workers_shared": 13,
    "workers_solo": 2,
    "hashrate": 6067885653,
    "hashrate_shared": 5099901433,
    "hashrate_solo": 967984220,
    "network_hashrate": 6957883569691.688,
    "difficulty": 97200,
    "estimate": "0.03598",
    "24h_blocks": 1,
    "24h_blocks_shared": 1,
    "24h_blocks_solo": 0,
    "24h_btc": 0.00017371,
    "24h_btc_shared": 0.00017371,
    "24h_btc_solo": 0,
    "24h_coins": 63.12,
    "24h_coinsUSD": 0.72,
    "percent_blocks": 1,
    "lastblock": 15076768,
    "timesincelast": 1350,
    "lastblock_shared": 15076768,
    "timesincelast_shared": 1350,
    "lastblock_solo": 15068111,
    "timesincelast_solo": 130255,
    "site": "http://www.digibyte.io",
    "bitkointalk": "https://bitkointalk.org/index.php?topic=408268.0",
    "github": "https://github.com/digibyte/digibyte",
    "explorer": "https://digiexplorer.info",
    "discord": ""
  },
  "DGB-SHA": {
    "algo": "sha256",
    "port": 3333,
    "name": "Digibyte",
    "height": 15076861,
    "block_reward": "454.49173007",
    "fees": 0.9,
    "fees_solo": 0.9,
    "workers": 0,
    "workers_shared": 0,
    "workers_solo": 0,
    "hashrate": 0,
    "hashrate_shared": 0,
    "hashrate_solo": 0,
    "network_hashrate": 86985520958494030,
    "difficulty": 1215173690,
    "estimate": "0.00288",
    "24h_blocks": 0,
    "24h_blocks_shared": 0,
    "24h_blocks_solo": 0,
    "24h_btc": 0,
    "24h_btc_shared": 0,
    "24h_btc_solo": 0,
    "24h_coins": 0,
    "24h_coinsUSD": 0,
    "percent_blocks": 0,
    "lastblock": 0,
    "timesincelast": 0,
    "lastblock_shared": 0,
    "timesincelast_shared": 0,
    "lastblock_solo": 0,
    "timesincelast_solo": 0,
    "site": "http://www.digibyte.io",
    "bitkointalk": "https://bitkointalk.org/index.php?topic=408268.0",
    "github": "https://github.com/digibyte/digibyte",
    "explorer": "https://digiexplorer.info",
    "discord": ""
  },
  "LCC": {
    "algo": "sha256",
    "port": 3344,
    "name": "Litecoin Cash",
    "height": 2797011,
    "block_reward": "31.25",
    "fees": 0.9,
    "fees_solo": 0.9,
    "workers": 3,
    "workers_shared": 3,
    "workers_solo": 0,
    "hashrate": 0,
    "hashrate_shared": 0,
    "hashrate_solo": 0,
    "network_hashrate": 278924550398805.3,
    "difficulty": 20595190,
    "estimate": "0.00442",
    "24h_blocks": 0,
    "24h_blocks_shared": 0,
    "24h_blocks_solo": 0,
    "24h_btc": 0,
    "24h_btc_shared": 0,
    "24h_btc_solo": 0,
    "24h_coins": 0,
    "24h_coinsUSD": 0,
    "percent_blocks": 0,
    "lastblock": 2793746,
    "timesincelast": 257446,
    "lastblock_shared": 2791579,
    "timesincelast_shared": 422199,
    "lastblock_solo": 2793746,
    "timesincelast_solo": 257446,
    "site": "https://litecoincash.sh",
    "bitkointalk": "https://bitkointalk.org/index.php?topic=2876538",
    "github": "https://github.com/litecoincash-project",
    "explorer": "https://chainz.cryptoid.info/lcc",
    "discord": "https://discordapp.com/invite/F2nZXnW"
  }
}
```

Примечание: на дату 19.05.2022 криптовалюты RVN нет на данном пуле.

В данном фрагменте видно, что, кроме RVN, добывается еще множество других криптовалют: DGB, LCC и др.

Следует отметить, что на форуме есть раздел, посвященный новым криптовалютам, в котором есть англоязычная «ветка» [4], где, как правило, отражено описание всех криптовалют. Отслеживая появление новых сообщений в этой ветке – периодически опрашивая форум, можно находить информацию о новых пулах.

Таким образом, периодически опрашивая форум, можно составить актуальный список пулов, в который будут оперативно заноситься также малоизвестные пулы – расширяющие возможности дальнейшего поиска новых криптовалют.

Периодически опрашивая обновляемый список пулов, можно получать список криптовалют и находить в нем новые, только появившиеся криптовалюты.

На 20.05.2022 был обновлен список всех найденных на форуме пулов. В списке указаны первые 8 адресов пулов и количество криптовалют в которых они упоминаются:

- 1) gos.cx – 157;
- 2) bsod.pw – 155;
- 3) arcpool.com – 107;
- 4) acpool.pro – 93;
- 5) lycheebit.com – 86;
- 6) eu1.arcpool.com – 82;
- 7) youcrazy.me – 81;
- 8) 0769.it – 79

Для наглядности построен график (рис. 11), на котором указаны найденные пулы и количество криптовалют, в которых они упоминаются. Для удобства отображения указаны только пулы с 25 и большим количеством упоминаний.

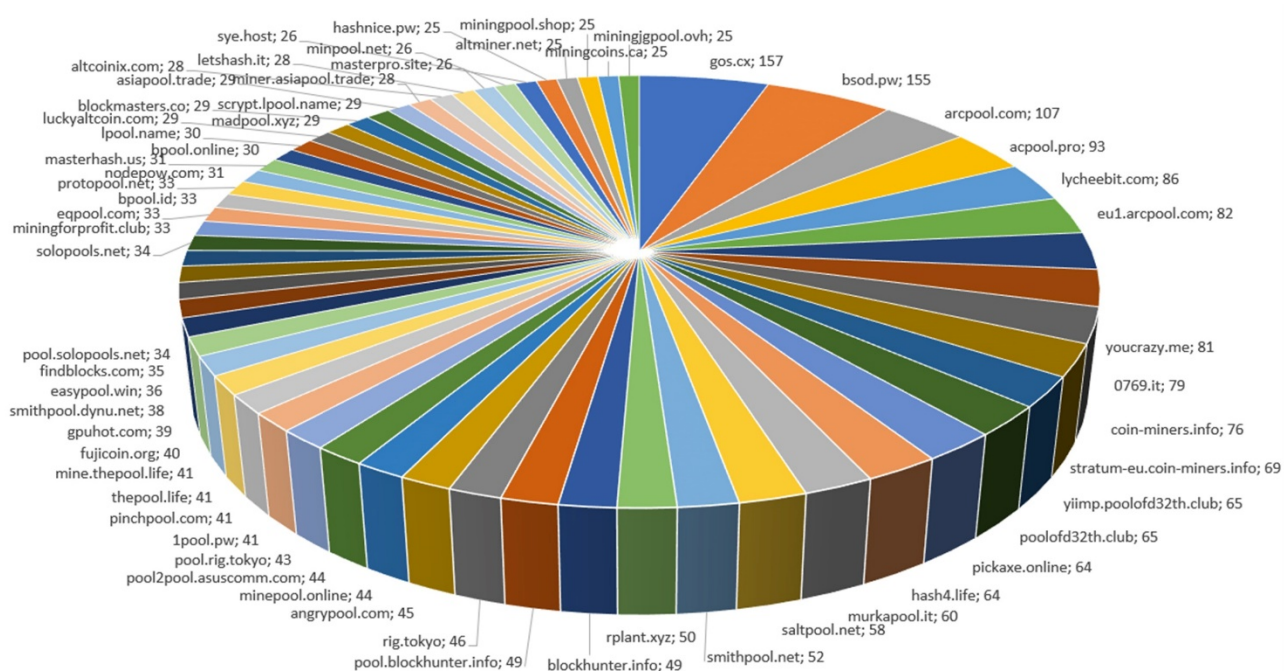


Рис. 11. Список пулов, найденных на форуме Bitkointalk, а также частота их упоминания в описании криптовалют

Для примера: используя предложенный алгоритм, на малоизвестном пуле wspool.zarpto.org, через рассмотренный в статье API была найдена криптовалюта SNARK, которая не отображалась (20.05.2022) в списке новых монет на популярном сервисе mining-poolstat.stream, что было продемонстрировано в выступлении. Эта криптовалюта появилась в списке новых криптовалют только вечером этого дня.

Таким образом, криптовалютный рынок очень подвижен, появляются новые криптовалюты и применение нового подхода к поиску новых майнинг-пулов приводит к повышению скорости нахождения новых криптовалют. Эффективность предложенного решения подтверждается нахождением новой криптовалюты (например, SNARK) раньше, чем она появилась на популярном сервисе поиска новых криптовалют.

Список источников

1. Сложность сети // Блокчейн.ком Самый надежный криптокошелек и биржа. URL: <https://www.blockchain.com/ru/charts/difficulty> (дата обращения: 20.05.2022).
2. Распределение хэшрейта // Блокчейн.ком | Самый надежный криптокошелек и биржа. URL: <https://www.blockchain.com/charts/pools> (дата обращения: 20.05.2022).
3. Home | Ravencoin. URL: <https://ravencoin.org> (дата обращения: 20.05.2022).
4. Announcements (Altcoins) / Bitcoin Forum. URL: <https://bitcointalk.org/index.php?board=159.0> (дата обращения: 20.05.2022).
5. [ANN] Ravencoin [RVN] PoW GPU Mining | Asset Transfer Blockchain (Updated ANN) / Bitcoin Forum. URL: <https://bitcointalk.org/index.php?topic=3238497> (дата обращения: 20.05.2022).
6. Роскомнадзор частично заблокировал криптовалютный форум BitcoinTalk // Биткоин, криптовалюты, блокчейн, DeFi – Bits Media. URL: <https://bits.media/roskomnadzor-chastichno-zablokiroval-kriptovalyutnyy-forum-bitcointalk/> (дата обращения: 20.05.2022).
7. Cryptocurrency Prices, Charts And Market Capitalizations | CoinMarketCap. URL: <https://coinmarketcap.com> (дата обращения: 20.05.2022).
8. Ravencoin (RVN) Цена, Графики, Рыночная капитализация // Cryptocurrency Prices, Charts And Market Capitalizations | CoinMarketCap. URL: <https://coinmarketcap.com/ru/currencies/ravencoin> (дата обращения: 20.05.2022).
9. New PoW Coins / Mining Pool Stats. URL: <https://miningpoolstats.stream/newcoins> (дата обращения: 20.05.2022).
10. Gos.cx: Best Alternative Crypto Mining Pool. URL: <https://gos.cx> (дата обращения: 20.05.2022).
11. Gos.cx API // Gos.cx: Best Alternative Crypto Mining Pool. URL: <https://gos.cx/site/api> (дата обращения: 20.05.2022).
12. Tpruvot / yiimp / GitHub. URL: <https://github.com/tpruvot/yiimp> (дата обращения: 20.05.2022).

SEARCH FOR NEW MINING POOLS FOR OPERATIONAL TRACKING OF THE APPEARANCE OF NEW CRYPTOCURRENCIES

I.I. Barilo

Kostroma State University, Kostroma, Russian Federation

Abstract. The article outlines the relevant problem of finding new cryptocurrencies, discusses the basic concepts related to their production, existing methods for obtaining information, popular information services that publish information about cryptocurrencies, as well as ways to disseminate information about cryptocurrencies from their developers. An approach to the search for new mining pools is proposed; the stages of their automatic detection and the features of checking the correctness of their determination are considered; the possibilities of obtaining information from mining pools and their use to search for new cryptocurrencies are presented. A search was made for all pools using the proposed method, and a list of them was prepared indicating the frequency of their use in the extraction of various cryptocurrencies. A practical example of using the proposed approach is given, when a new cryptocurrency was found earlier than one of the popular services.

Keywords: cryptocurrency, pool, mining, regular expressions, search for new cryptocurrencies

ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ СИСТЕМАТИЧЕСКОГО АНАЛИЗА ИССЛЕДОВАНИЙ СОЦИАЛЬНЫХ ФЕНОМЕНОВ: ПРОБЛЕМА ОБЕСПЕЧЕНИЯ ГОМОГЕННОСТИ ЗНАЧЕНИЙ КОНЦЕПТОВ

А.А. Барышев, Г.Н. Сербина

Национальный исследовательский Томский государственный университет, Томск, Россия

Аннотация. Проблема обеспечения гомогенности значений концептов имеет решающее значение для составления обзоров исследований социальных феноменов. Целью статьи является оценка возможности существующих требований для составления систематических обзоров по социальным наукам. Рассматриваются типы обзоров научной литературы и специфика протоколов и стандартов систематического обзора. Показано, что проблема их применения для анализа исследований по социальной тематике связана с неоднозначностью социальных понятий. В этом контексте метод картирования значений социальных концептов в дискурсивных полях выдвигается как инструмент отбора и обеспечения однородности значений. С этой точки зрения можно заложить основу жанра систематического обзора в гуманитарных науках путем уточнения существующих стандартов составления обзоров в этой области. Выдвинутое в статье предложение подкреплено прототипами диахронической и синхронической карт рассеивания смыслов концепта благополучия.

Ключевые слова: социальный концепт, значение концепта, дискурс, дискурсивные поля, систематический обзор, практики благополучия, социальное картирование, библиографический анализ

Источник финансирования: исследование выполнено при поддержке Программы развития ТГУ «Приоритет-2030».

Введение: профессионализация деятельности по производству обзоров научных исследований и переход к ее автоматизации

Становление общества знания сопровождается углублением дифференциации научных исследований и стремительным ростом их числа. Все это ведет к резкому увеличению объема научной информации, публикуемой в научных журналах. В этих условиях со стороны исследователей, или *knowledge users* в широком смысле, формируется спрос на услуги так называемого «синтеза знания» (*knowledge synthesis*). В свою очередь, формируется и соответствующее предложение в виде *the art and science of knowledge synthesis* [1], проявляющееся в росте числа систематических обзоров литературы по различным направлениям исследований [2, 3].

Процесс институционализации этого вида научного труда происходит путем создания сообществ (в медицине, напр., Cochrane and the Joanna Briggs Institute (JBI)), издания специализированных журналов (Systematic Reviews) и руководств (Cochrane handbook), спецификации родовых свойств производимых «обзорно-аналитических продуктов» [4], их дальнейшей дифференциации [5], стандартизацией процесса их производства (стандарты PRISMA, SALSA, PSASAR и т.д.). Таким образом, наблюдается автономизация особого вида исследовательских практик со своими формализованными правилами, профессиональными коммуникациями и авторитетными представителями в отдельных дисциплинарных сферах, например, D. Hernández-Torrano [3, 6] – в психологии и образовании, I. Kavakiotis [7] – в области биотехнологий, J. Tomasi [8] – в химии и т.д.

В сферу этих новых практик вместе с общемировым трендом на расширение использования ИТ в науке происходит приоритетное проникновение автоматизированных технологий «понимания» текстов и «синтеза знаний», а именно: программ автоматизированного ревьюирования научной литературы (systematic review automation toolboxes/tools to support systematic reviews) [9, 10] на основе использования технологий обработки естественного языка, машинного обучения, искусственного интеллекта. Это ведет к появлению формальных и неформальных сообществ разработчиков программ для автоматизации работ по составлению обзоров [11] и, соответственно, конкуренции между ними [12].

Типы обзоров научной литературы и специфика систематического ревью

При обращении к вопросу об уже существующих типах обзоров научной литературы стоит отметить работу Grant, M.J. и Booth, A., которые выделяют 14 типов обзоров научной литературы в зависимости от целей и фокусов анализа: критический обзор (critical review), литературный обзор (literature review), *картирующее ревью* / систематическая карта / (mapping review/systematic map), мета-анализ (meta-analysis), обзор смешанных исследований / обзор смешанных методов (mixed studies review / mixed methods review), (обзор / overview), качественный систематический обзор / обобщение качественных данных (qualitative systematic review/qualitative evidence synthesis), быстрый обзор (rapid review), предварительный обзор (scoping review), современный обзор (state-of-the-art review), систематический обзор (systematic review), систематический поиск и обзор (systematic search and review), систематизированный обзор (systematized review), зонтичный обзор (umbrella review) [5].

Систематический обзор определяется как «наиболее известный тип обзора, направленный на систематический поиск, оценку и обобщение научных данных, часто придерживаясь рекомендаций по проведению обзора, разрабатываемых Cochrane Collaboration или NHS Centre for Reviews and Dissemination». Такой вид обзора «прозрачен в отчетности о своих методах с целью помочь другим воспроизвести процесс» [5, с. 95]. Основными характеристиками систематического обзора являются: 1) четко поставленный исследовательский вопрос, на который должно ответить исследование; 2) наличие четко сформулированных целей, которые имеют явный и воспроизводимый метод; 3) перечисление условий поиска, включающих все связанные исследования, отвечающие критериям приемлемости; 4) оценка качества/валидности выбранных исследований; 5) систематическое представление и синтез данных, извлеченных из выбранных исследований; 6) обеспечение доступности результатов исследования для научных целей и принятия решений.

Для соблюдения данных характеристик систематических обзоров разрабатываются специальные протоколы, то есть наборы стандартизированных правил для его качественного проведения в соответствии с научными требованиями.

Протоколы систематического обзора и проблема их применения для анализа исследований по социальной тематике

Подобные протоколы чаще всего разрабатываются профессиональными организациями, такими как Cochrane Systematic Reviews [13], IOM Standards for Initiating a Systematic Review [14], NHS Centre for Reviews and Dissemination [15].

Cochrane Systematic Review и IOM Standards применяются в области систематического и мета-анализа медицинских исследований. PSALSAR был разработан для области исследований об окружающей среде [16], но может применяться шире. Рассмотрим его в качестве примера исследовательского протокола. Данный протокол включает 7 шагов, как следует из названия: протокол (Protocol), поиск (Search), оценка (Appraisal), синтез (Synthesis), анализ (Analysis), отчет (Report). Каждый из шагов протоколирует правила, по которым разворачивается систематический обзор: от общей концепции исследований до написания отчета.

Не следует также путать протоколы проведения систематического обзора с шаблонами и рамками различных этапов проведения обзора. Например, шаблоны для постановки исследовательской цели систематического обзора (PICOS и т.д.) или шаблоны отчетности (PRISMA [17] и т.д.).

Таким образом, для всех этапов проведения систематического обзора в естественных и медицинских науках существуют строгие протокольные процедуры, что обеспечивает качество и воспроизводимость результатов систематических обзоров.

Обзоры исследований по социальной проблематике традиционно представляли собой особый жанр научного и литературного творчества. Многие выдающиеся произведения в об-

ласти социальных наук возникали именно как авторские критические обзоры исследований современников и предшественников. Таковы, например, труды К. Маркса, М. Фуко, в которых «список использованной литературы» может охватывать не одну сотню источников. И эти источники надежно «закрывают» все основные аспекты изучаемого явления и направления его исследования на достаточно долгую перспективу. Однако в настоящее время ситуация принципиально иная. Так, например, поиск в Scopus по ключевому слову *well being* и области знаний *social science* выдает 56 813 результатов за 2020 г., 52 568 за 2019, 47 753 за 2018, 43 609 за 2017, 40 807 за 2016, 37 652 за 2015 г. То есть за 5 лет количество исследований ежегодно увеличивалось в 1,5 раза (рис. 1).

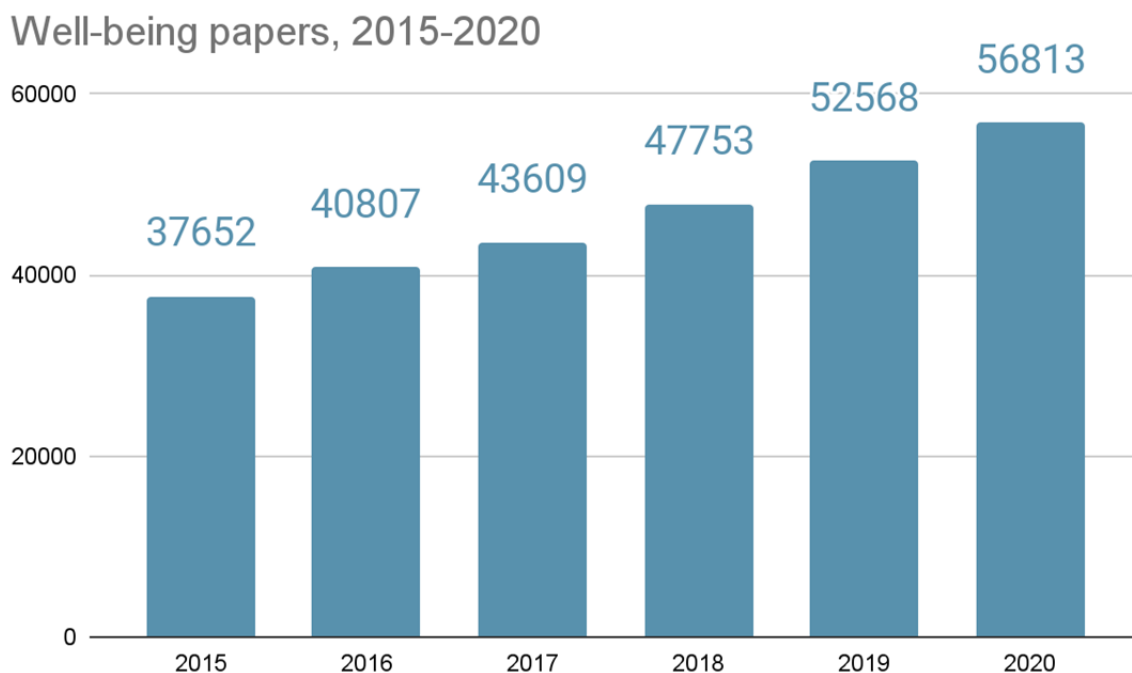


Рис. 1. Поле исследований благополучия в Scopus по запросу well-being, направление Social Science

Очевидно, что за этим количественным ростом скрываются новые аспекты феномена благополучия и новые направления и подходы к его изучению. Но не это главное: одновременно с ростом числа исследований и публикаций по теме благополучия нарастают разночтения в трактовке самого понятия [18, с. 3], появляются новые значения этого термина.

Существующими руководствами по систематическому анализу литературы рекомендуется проводить уточнение концептов посредством процедуры предварительного анализа концептов [19], соответствующего этапу «Appraisal» протокола PSALSAR. Однако если применительно к техническим и естественным наукам уточнение понятия состоит в обеспечении максимально возможной на данный момент универсальности его определения путем отвлечения от локальных условий [21, с. 46–49], что представляет собой по большей части *техническую* задачу, то для наук об обществе уточнение редко достигается посредством более строгих формально-логических обобщений [21, с. 17].

Это объясняется тем, что определение содержания социальных концептов (вновь обратимся к термину «благополучие») является процессом социальным (правильнее даже сказать «*политическим*» [22, с. 114; 23]), а не просто теоретическим, или узкодисциплинарным. Участниками этого процесса являются как сами «люди», действия которых подлежат концептуализации, так и ученые, политики, законодатели, представители контрольных, регистрирующих органов, органов масс-медиа и т.д. В процессе интеракций всех этих акторов происходит конструирование благополучия как социального феномена. В результате содержание концепта «благополучие» задается скорее полем распределения его значений, чем его однозначным определением (см., например, [24, с. 155]). Именно поэтому простого «уточне-

ния» понятий в процессе подготовки к составлению систематического ревью по социальной проблематике оказывается недостаточно, хотя многие современные обзоры, связанные в том числе и с понятием благополучия, вполне справляются со своими задачами и без пересмотра методологических приемов протокола качественного анализа концептов (см., например, [25–31]).

Это оказывается возможным и вполне оправданным по причине того, что авторы указанных обзоров могут полностью полагаться в процессе их составления на сложившуюся медицинскую традицию: благодаря выбору узкой проблемы процедура качественной оценки обеспечения гомогенности значений концепта выполняется практически на интуитивном уровне, поскольку благополучие в приведенных в сноске статьях имеет заведомо одинаковое значение создания / нарушения функциональной дееспособности индивида при использовании определенных гаджетов в случаях вовлеченности в некоторые практики или при работе в каких-либо особых условиях. При выполнении же систематического обзора по собственно социальной проблеме, скажем, «изменения благополучия жителей / населения / конкретной группы на территории X в результате проведения мероприятий Y в условиях Z» возникают сложности с соблюдением требований протоколов по части обеспечения гомогенности значений концептов из разных источников, включаемых в ревью. Это связано с тем, что «благополучие» в этом случае может «плавать», например, между его объективным и субъективным значениями, между его пониманиями в терминах способностей, с одной стороны, и возможностей человека – с другой, и между различными практическими смыслами способов его созидания: путем патерналистского или принудительного внедрения наиболее благоприятных изменений в образах жизни людей, путем самостоятельного культивирования индивидами «заботы о себе», посредством нарративной работы над идентичностью или формирования опыта сопротивления навязываемым со стороны общества стандартам «хорошей жизни» (англ. *good life*) [32, 33].

Смешение в одном обзоре этих разнообразных смыслов, особенно при автоматизированном отборе исследований, неизбежно приведет к размыванию его ценности для будущих пользователей. Соответственно, фундаментальной проблемой при разработке методологии составления систематических обзоров (тем более с использованием возможностей искусственного интеллекта) является проблема распознавания гомогенных значений социальных концептов.

Картирование полей распределения значений социальных концептов как инструмент селекции их гомогенных значений

Чтобы решить проблему распознавания и селекции для целей составления систематических обзоров гомогенных значений социальных концептов, необходимо обеспечить их представление в рамках некоторой единой таксономии. При этом следует иметь в виду то, что поскольку любая система значений представляет собой нестабильное образование, постоянно изменяющееся вследствие изменения состава акторов и их позиций по отношению к конструируемому феномену, постольку «единая таксономия» может быть надежно зафиксирована лишь «на данный момент». Это значит, что процедура «расознавания и селекции» должна стать протокольной нормой при составлении каждого обзора.

До недавних пор способом представления репрезентации значений социальных концептов/феноменов являлся метод социальной картографии, или «феноменографии» [34–36]. Последнее время под «картированием» (*mapping*) понимается построение сетевой визуализации (*map*) библиографических данных обзриваемых исследований [3], а также (и все чаще) различных аспектов взаимодействий акторов в социальных сетях [37, 38]. Очевидно, что подобные визуализации служат в качестве изображения («картины») сетевого объекта, но никак не его «карты», дающей представление о его значимой (в соответствии с необходимыми для этого легендой и координатной разметкой) дифференциации.

Недостатком же методов *Social Cartography* является то, что концептуальное поле «отрисовывается» в *ad hoc* выбранных координатах, например, “Support/Opposition –Individual/Social” [39], или “Emergent/Structured – Consensual/Contested” [40] и т.п. (см. также [41, 42]).

Что же касается упомянутого выше «картирующего ревью», то под «систематической картой» подразумевается схематическое изображение тех аспектов обозреваемого феномена, которые на данный момент уже исследованы [43], чтобы выявить «пробелы в исследованиях, где необходимы дальнейшие первичные исследования» [44]. Очевидно, что характерное для обзора данного типа картирование «аспектов» имеет мало общего с необходимой для спецификации социальных концептов гибкой таксономией их конкретных значений. Производство же новых значений социальных концептов может прекратиться только с их уходом из социальной жизни. Так, в отношении благополучия можно констатировать продолжающееся формирование его значений, связанных с процессами самосозидания человеческой индивидуальности, с противостоянием (англ. Resistance [45], counter-conduct, art of not being governed [46]); человека паттернам достойной жизни, выступающим на деле мягкими инструментами ее говернментализации. Термин М. Фуко «говернментальность» означает новый тип власти, которая добивается добровольности подчинения человека, его респонсификации по отношению к социальному порядку [47], становящемуся вместо системы силового принуждения источником всего необходимого для получения жизненного удовлетворения (англ. life satisfaction) [48].

Из данного неструктурированного мини-обзора видно, что генератором новых значений понятия благополучия служит напряженность между восприятиями счастливой жизни и жизнью осмысленной (англ. between a happy life and a meaningful one [49]).

Поскольку наличие напряженностей в социальных феноменах выступает причиной их гибкости и изменчивости, постольку основные линии напряженности и должны выступать в качестве координатных осей при картировании разнообразия значений выражающих их концептов. Однако в каком дискурсивном поле реализуется это разнообразие? Выдвинем предположение о том, что это поле, связывающее различные значения, задается общей ключевой метафорой, что необходимо, в том числе, и для обеспечения взаимопонимания исследователей при работе с таким сложным междисциплинарным объектом, каким в нашем случае является благополучие. Выделяются три ключевые метафоры, во-первых, представляющие объект в целом и, во-вторых, позволяющие благодаря своему доконцептуальному характеру наладить взаимодействие ученых различных специальностей.

На первом этапе благополучие определяется как состояние, выступающее функцией социальной системы. Этому функциональному состоянию свойственна напряженность по линии интенциональности, задаваемой противоположностью определяющих предрасположенности людей к счастливой жизни полюсов («потребности» / «способности») и по линии способов реализации этой интенциональности (от простой адаптации к существующим условиям до их кардинальных изменений). В дальнейшем благополучие представляется как (социальный и экономический) процесс со свойственной ему напряженностью между линиями ценностей (от экономических до социальных) и линией соответствующих этим ценностям индивидуальных мотивов (от объективных до субъективных).

Третьей метафорой благополучия выступает практика. В данной статье попытаемся дать координатную разметку для картирования исследований практик благополучия. Метафорика практики погружает исследование в атмосферу повседневности, жизненного мира человека, работы его практического интеллекта. Степень напряженности, которую испытывает феномен благополучия, здесь достигает своего максимума. Это проявляется в том, что любая категоризация практик выступает крайне ненадежной, требует от актора и исследователя большого напряжения для удержания фокусировки на ней без срыва в иные категоризации.

Обратимся к основным линиям напряженности, в которых происходит повседневное конструирование практик благополучия. По первой линии та или иная практика получает направленность либо на самого человека, либо на других как на причины беспокойства или заботы. Условно «концы» этой линии можно обозначить как «политики» (это – про «других») и «искусства» (когда, например, практикуется «забота о себе»). Вторая линия разделяет практики по источникам формирования их смысла. Смыслы возникают либо как результаты

нарративной работы, в которой создается определенная идентичность, либо прямым «воплощённым смыслообразованием» (англ. enacted sense making) [50], то есть путем перформативной работы как противоположности нарративной рефлексивности [51]. Условно данная напряженность оказывается зажатой между «нарративами» и «перформативами». В результате макет карты анализа практик благополучия принимает следующий вид (рис. 2).



Рис. 2. Картография дискурсивного поля практик благополучия

Данный вариант картирования дискурса благополучия, когда анализируется его часть, задаваемая отдельной метафорой, с определенной долей условности можно назвать диахронным картированием. Соответственно, представление дискурса одновременно во всех метафорах целесообразно обозначить как синхронное картирование (рис. 3).

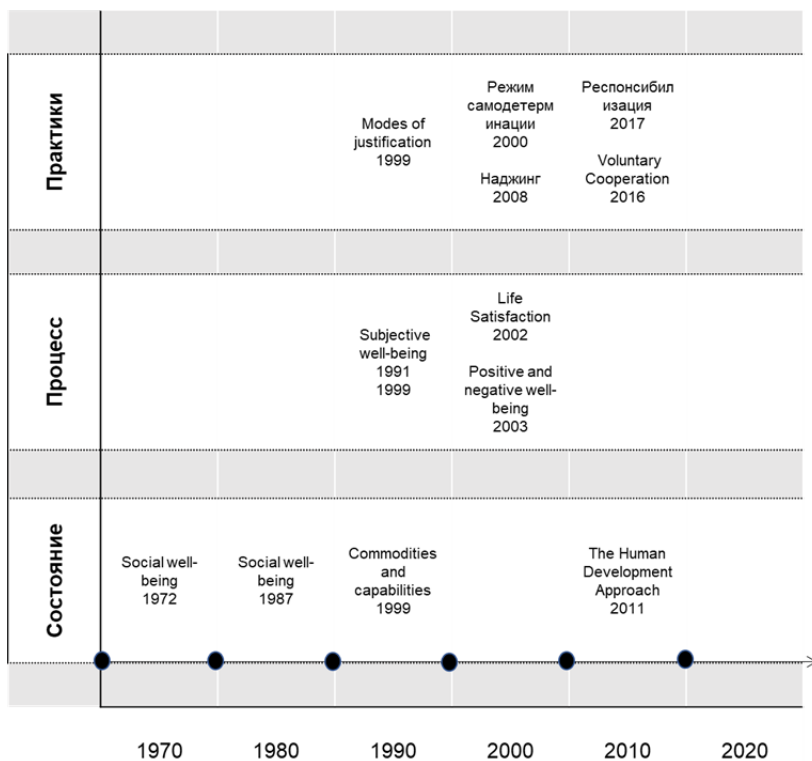


Рис. 3. Макет синхронного картирования поля исследований благополучия

Заклучение

Проведенный анализ типов, стандартных процедур систематического ревьюирования показывает применимость данного вида научных работ к анализу исследований социальных феноменов при определенной доработке существующих стандартов составления обзоров в части требований, относящихся к оценке концептов и обеспечения их однородности.

Предложенный в данной статье абрис методологии картирования рассеивания значений социальных концептов может быть развит до уровня рабочего инструмента селекции их гомогенных значений для целей составления систематических обзоров в области социальных наук. При диахронном картировании появляется возможность выявления крупных различий (на уровне ключевых метафор) в употреблении социальных понятий, проявляющихся на всем протяжении их жизненного цикла. При синхронном картировании обеспечивается выделение конкретных значений, пригодных для применения автоматизированных технологий в процессе составления обзоров.

Отработку предлагаемых макетов карт социальных концептов, наполнение их библиографическими сетями слов и авторов предполагается осуществить на понятиях предпринимательства, благотворительности, благополучия и политического экстремизма в рамках Программы развития ТГУ («Приоритет-2030»).

Полученные карты дискурсивных полей с наполнением из библиографических сетей слов, связанных с термином «благополучие», и авторов исследований благополучия могут служить инструментом селекции гомогенных значений этого концепта.

Список источников

1. Tricco A.C., Tetzlaff J., Moher D. The art and science of knowledge synthesis // *Journal of Clinical Epidemiology*. 2011. № 64 (1). P. 11–20.
2. Greenhalgh, C., Montgomery, P. A systematic review of the barriers to and facilitators of the use of evidence by philanthropists when determining which charities (including health charities or programmes) to fund // *Systematic Review*. 2020. № 199 (9). <https://doi.org/10.1186/s13643-020-01448-w>
3. Hernández-Torrano D., Ibrayeva L., Sparks J., Lim N., Clementi A., Almukhambetova A., Nurtayev Y. and Muratkyzy A. Mental Health and Well-Being of University Students: A Bibliometric Mapping of the Literature // *Frontiers in Psychology*. 2020. 11:1226. doi: 10.3389/fpsyg.2020.01226
4. Лаврик, О. Л., Калюжная, Т. А., Плешакова, М. А. Систематический обзор как вид обзорно-аналитических продуктов // *Библиосфера*. 2019. № 2. С. 33–51.
5. Grant M.J., Booth A. A typology of reviews: an analysis of 14 review types // *Health Information & Libraries Journal*. 2009. № 26 (2). P. 91–108.
6. Hernández-Torrano D., Ibrayeva L. Creativity and education: A bibliometric mapping of the research literature (1975–2019) // *Thinking skills and creativity*. 2020. № 35 (100625). <https://doi.org/10.1016/j.tsc.2019.100625>
7. Kavakiotis I., Tsave O., Salifoglou A., Maglaveras N., Vlahavas I., Chouvarda I. Machine Learning and Data Mining Methods in Diabetes Research // *Computational and Structural Biotechnology Journal*. 2017. Vol. 15. P. 104–116. <https://doi.org/10.1016/j.csbj.2016.12.005>.
8. Tomasi J., Mennucci B., Cammi R. Quantum mechanical continuum solvation models // *Chemical Reviews*. 2005. № 105 (8). P. 2999–3093. doi: 10.1021/cr9904009
9. Marshall I.J., Wallace B.C. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis // *Systematic reviews*. 2019. № 8 (163). <https://doi.org/10.1186/s13643-019-1074-9>
10. Marshall C., Brereton P. Systematic review toolbox: a catalogue of tools to support systematic reviews // *EASE '15: Proceedings of the 19th International Conference on Evaluation and Assessment in Software Engineering*. 2015, April. P. 1–6. <https://doi.org/10.1145/2745802.2745827>
11. Rayyan – faster systematic reviews. URL: <https://www.rayyan.ai/> (дата обращения: 01.05.2022).
12. Robledo S., Grisales A., Hughes M., Eggers F. “Hasta la vista, baby”– will machine learning terminate human literature reviews in entrepreneurship? // *Journal of Small Business Management*. 2021. P. 1–30.
13. Henderson L.K., Craig J.C., Willis N.S., Tovey D. Webster A.C. How to write a Cochrane systematic review // *Nephrology*. 2010. № 15. P. 617–624. <https://doi.org/10.1111/j.1440-1797.2010.01380.x>
14. Institute of Medicine (US) Committee on Standards for Systematic Reviews of Comparative Effectiveness Research. Finding What Works in Health Care: Standards for Systematic Reviews. Washington (DC) : National Academies Press (US), 2011. URL: https://www.nihlibrary.nih.gov/sites/default/files/IOM_Standards_for_Systematic_Review_2010_Insert.pdf (дата обращения: 01.05.2022).
15. Systematic Reviews // Centre for Reviews and Dissemination. URL: <https://www.york.ac.uk/crd/guidance/> (дата обращения: 01.05.2022).

16. Mengist W., Soromessa T., Legese G. Method for conducting systematic literature review and meta-analysis for environmental science research // *MethodsX*. 2020. № 7. P. 100777.
17. Liberati A., Altman D.G., Tetzlaff J., Mulrow C., Gøtzsche P.C., Ioannidis J.P., Clarke M., Devereaux P.J., Kleijnen J., Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration // *PLoS Med*. 2009. Vol. 21, № 6(7). doi: 10.1371/journal.pmed.1000100
18. White S.C. Analysing wellbeing: a framework for development practice // *Development in practice*. 2010. № 20 (2). P. 158–172.
19. Munn Z., Peters M.D.J., Stern C. et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach // *BMC Med Res Methodol*. 2018. Vol. 18, № 143. <https://doi.org/10.1186/s12874-018-0611-x>
20. Дойч Д. Начало бесконечности: Объяснения, которые меняют мир. М. : Альпина нон-фикшн, 2016. С. 46–49.
21. Барышева Г.А., Недоспасова О.П., Фролова Е.А., Маланина В.А., Рождественская Е.М., Барышев А.А. Теория и практика социального предпринимательства. Томск : STT Publishing, 2018.
22. Dey P., Steyaert C. Myth in social entrepreneurship research: an inquiry into rationalist, ideological and dialectic practices of de mystification // *Social Entrepreneurship. An Affirmative Critique*. Cheltenham, UK : Edward Elgar Publishing Ltd., 2018. P. 114.
23. Бикбов А.Т. Грамматика порядка: Историческая социология понятий, которые меняют нашу реальность. М. : НИУ «Высшая школа экономики», 2014. 432 с.
24. Baryshev A. The regional spectra for the entrepreneurial motives and the dissemination of entrepreneurs' knowledge in the context of modern discursive formation of entrepreneurship // *Economic and Social Development SPECIAL EDITION XIV International Conference Russian Regions in the Focus of Changes Book of Proceedings*. 2019. P. 155.
25. Kozusznik M.W., Maricutoiu L.P., Peiró J.M., Vîrgă D.M., Soriano A., Mateo-Cecilia C. Decoupling office energy efficiency from employees' well-being and performance: a systematic review // *Frontiers in psychology*. 2019. № 10 (293).
26. Cornet V.P., Holden R.J. Systematic review of smartphone-based passive sensing for health and wellbeing // *Journal of biomedical informatics*. 2018. № 77. P. 120–132.
27. Kachouie R., Sedighadeli S., Abkenar A.B. The role of socially assistive robots in elderly wellbeing: A systematic review // *International Conference on Cross-Cultural Design*. 2017. P. 669–682.
28. Johnson D., Deterding S., Kuhn K.A., Staneva A., Stoyanov S., Hides L. Gamification for health and wellbeing: A systematic review of the literature // *Internet interventions*. 2016. № 6. P. 89–106.
29. Daykin N., Mansfield L., Meads C., Julier G., Tomlinson A., Payne A., Victo, C. What works for wellbeing? A systematic review of wellbeing outcomes for music and singing in adults // *Perspectives in public health*. 2018. № 138 (1). P. 39–46.
30. Best P., Manktelow R., Taylor B. Online communication, social media and adolescent wellbeing: A systematic narrative review // *Children and Youth Services Review*. 2014. № 41. P. 27–36.
31. Gestos M., Smith-Merry J., Campbell A. Representation of women in video games: A systematic review of literature in consideration of adult female wellbeing // *Cyberpsychology, Behavior, and Social Networking*. 2018. № 21 (9). P. 535–541.
32. Park N., Peterson C. Achieving and sustaining a good life // *Perspectives on Psychological Science*. 2009. № 4 (4). P. 422–428.
33. Щекотин Е.В. Проблема благополучия в турбулентном социуме: аспект безопасности // *Векторы благополучия: экономика и социум*. 2017. № 4 (27). С. 74–83.
34. Paulston R.G. A Spatial Turn in Comparative Education? Constructing a Social Cartography of Difference // *Discourse Formation in Comparative Education* / ed. by J. Schriewer. Frankfurt am Main : Peter Lang, 2003. 366 p.
35. Social cartography: Mapping ways of seeing social and educational change / ed. by R.G. Paulston. New York : Garland, 1996.
36. Marton F. Phenomenography: Exploring different conceptions of reality // *Qualitative Approaches to Evaluation in Education*. New York : Praeger, 1988. P. 176–205.
37. Bykov I.A., Martyanov D.S. Studying political communities in vk.com with network analysis // *Galactica Media: Journal of Media Studies*. 2021. № 3 (1). P. 64–78.
38. Mapping White Identity Terrorism and Racially or Ethnically Motivated Violent Extremism: A Social Network Analysis of Online Activity / Heather W.J., Matthews L.J., Moore P., DeNardo M.A., Marrone J.V., Jackson B.A., Marcellino W., Helmus T.C. Santa Monica, CA : RAND Corporation, 2022. URL: https://www.rand.org/pubs/research_reports/RRA1841-1.html (дата обращения: 01.05.2022).
39. Erkkila K. Ways of Seeing the Entrepreneurial Education Debates in the United States, the United Kingdom and Finland. New York : Garland, 1999. 19 p.
40. Snow D.A. Elaborating the discursive contexts of framing: Discursive fields and spaces // *Studies in symbolic interaction*. 2008. № 30. P. 3–28.

41. Dean G., Bell P., Vakhitova Z. Right-wing extremism in Australia: the rise of the new radical right // *Journal of Policing, Intelligence and Counter Terrorism*. 2016. № 11 (2). P. 121–142.
42. Ferreira J.J., Fernandes C.I., Kraus S. Entrepreneurship research: mapping intellectual structures and research trends // *Review of Managerial Science*. 2019. № 13 (1). P. 181–205.
43. Fernandez A, Black J, Jones M, Wilson L, Salvador-Carulla L, Astell-Burt T, Black D. Flooding and mental health: a systematic mapping review // *PLoS One*. 2015. Vol. 10, № 10 (4). P. e0119929. doi: 10.1371/journal.pone.0119929
44. Clapton J., Rutter D., Shariff N. SCIE Systematic mapping guidance UK 2009 // Social Care Institute for Excellence (SCIE). URL: <http://www.scie.org.uk/publications/researchresources/rr03.pdf> (дата обращения: 01.05.2022).
45. Armstrong N., Murphy E. Conceptualizing resistance // *Health*. 2012. № 16 (3). P. 314–326.
46. Lorenzini D. From counter-conduct to critical attitude: Michel Foucault and the art of not being governed quite so much // *Foucault Studies*. 2016. P. 7–21.
47. Siltaoja M., Malin V., Pyykkönen M. ‘We are all responsible now’: Governmentality and responsabilized subjects in corporate social responsibility // *Management Learning*. 2015. № 46 (4). P. 444–460.
48. McKay F. Psychocapital and Shangri-Las: How happiness became both a means and end to governmentality // *Health, Culture and Society*. 2013. № 5 (1). P. 36–50.
49. Moore K. *Understanding Wellbeing // Wellbeing and Aspirational Culture*. Palgrave Macmillan: Cham, 2019. P. 71–105.
50. Морган Г. Парадигмы, метафоры и решение задач в теории организации // *Теория организации : хрестоматия*. СПб. : Высшая школа менеджмента, 2010. С. 2–26.
51. Мелякова Ю., Жданенко С., Кальницький Е. От биополитики к прекарности коммодифицированных тел, или апогей перформанса // *Вісник НЮУ імені Ярослава Мудрого. Серія: Філософія, філософія права, політологія, соціологія*. 2021. № 4 (51). С. 11–33.

**PROSPECTS AND OPPORTUNITIES FOR SYSTEMATIC ANALYSIS OF STUDIES
ON SOCIAL PHENOMENA: THE PROBLEM OF ENSURING THE HOMOGENEITY
OF CONCEPTS’ MEANINGS**

A.A. Baryshev, G.N. Serbina

National Research Tomsk State University, Tomsk, Russian Federation

Abstract. The problem of ensuring the homogeneity of concepts’ meanings is crucial for reviewing of journal articles on social phenomena. The aim of the paper is to assess the feasibility of established standards for creating systematic reviews of literature on social issues. The types of scientific literature reviews and the specifics of the protocols and standards for a systematic review are considered. It is shown that the problem of their application for the analysis of research on social topics is connected with social concepts’ ambiguity. In this context, the method of the cartography of social concepts’ meanings in discursive fields is hypothesized as a tool for selecting and ensuring meanings’ homogeneity. From this perspective, it is possible to provide the basis for the genre of systematic review in the humanities by refinement of existing standards for compiling reviews in this field. The proposal put forward in the paper is supported by the prototypes of diachronic and synchronic maps of the meanings’ dispersion for the concept of well-being.

Keywords: social concept, concept’s meaning, discourse, discursive field, systematic review, well-being practices, social cartography, bibliographic analysis

СЕЛЬСКАЯ ШКОЛА: АНАЛИЗ ПРЕДСТАВЛЕНИЯ ПРОБЛЕМ НА ПУБЛИЧНЫХ АРЕНАХ

Е.С. Беляева, И.Э. Петрова

*Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского,
Нижний Новгород, Россия*

Аннотация. Тематика сохранения и развития сельских школ в современной России является злободневной, но представление проблем сельских учителей, значимости школы для сельского социума происходит в основном в рамках работы органов государственной власти и учреждений ведомственного подчинения сферы образования. Участие в этом процессе педагогов и жителей села (родителей школьников и неравнодушных активистов) затруднено в силу ограниченного доступа. Исследование посвящено цели выяснения акторов и способов конструирования мнений о проблемах сельской школы на разных публичных аренах, в том числе в социальных сетях (страницы сообществ, посвященных поддержке сельских инициатив). Сделаны выводы о недостаточной активности сельских педагогов в продвижении позиции школьного сообщества, о ресурсах и барьерах закрепления кадров системы образования на селе.

Ключевые слова: сельские школы, педагоги, сельский социум, конструирование проблем, публичные арены, социальные сети

Конструирование социальной проблемы закрепления педагогических кадров в сельских школах происходит на официальных, в особенности ведомственных, аренах. Проблема учителей сельских школ не находит полномасштабного решения в связи с разрывом целеполагания на разных уровнях принятия управленческих решений, а также с неоднозначной локализацией носителя этой проблемы. Выдвижение утверждений-требований происходит на уровне Правительства РФ и Министерства Просвещения РФ, но при этом мало учитывается специфика регионов; далеко не все губернаторы и правительства субъектов РФ обращают пристальное внимание на условия работы сельских учителей. Тогда проблема концентрируется на уровне директоров школ, и даже не все педагогические работники, тем более молодые учителя, готовы продвигать социальную проблему закрепления специалистов на публичных аренах в ущерб решению собственных личных и семейных социально-психологических задач.

Если школа на селе продолжает играть значимую роль, то более точно было бы связывать конструирование проблем закрепления педагогов с интересом сельского социума сохранить школу в конкретном сельском поселении. Истории о постройке новой школы или сохранении действующей в противовес оптимизации раз за разом возникают в лентах новостных агентств. Структуры гражданского общества на селе могут быть гораздо более эффективны в продвижении социальной проблемы развития сельской школы на публичных аренах и СМИ, потому что не связаны ведомственной иерархией, привлекают ресурсы других акторов и реализуют интерес именно носителей проблемы.

Парламентские слушания являются организационной формой обсуждения парламентом РФ значимых вопросов политики с привлечением должностных лиц, экспертов и общественности. Парламентские слушания могут рассматриваться как способ общественного обсуждения и участия общественности в законодательном процессе. Таким образом, на этой публичной арене могут быть представлены мнения разных вовлеченных акторов. По итогам слушаний могут быть приняты заключения и сформированы рекомендации [1].

Парламентские слушания 05 апреля 2018 г. «О состоянии и перспективах развития сельской школы в Российской Федерации» [1] собрали совершенно разных неравнодушных людей – представителей практически всех вовлеченных сторон социальной проблемы развития сельской школы. К сожалению, идея строительства социокультурных комплексов в сельской глубинке на основе школ не нашла воплощения, и только отдельные регионы по своей

инициативе продолжают этот интересный эксперимент. Высказанные замечания и предложения отражают реальное положение дел и интересы всех сторон: учителей, директоров, представителей общественных движений, экономистов, законодателей и ведомственных работников программ закрепления педагогических работников на селе. Общественное обсуждение позволяет конструировать социальную проблему именно ее носителям и представить максимальное число вовлеченных акторов.

Министр просвещения на парламентских слушаниях по вопросам качества школьного образования в 2019 г. отметила недостаток кадров по стране в целом (~10%) при условии, что учителя, в особенности в сельской местности, перегружены (в среднем более 1,7 ставки), что увеличивает нормативную потребность в кадрах практически вдвое [2].

Поправки в федеральный закон «Об образовании в Российской Федерации», позволяющие студентам старших курсов педагогических направлений колледжей и вузов работать в школе в качестве учителя и проводить кружки и секции был анонсирован на парламентских чтениях и введен в действие в 2020 г.

Дискуссии о сельских школах продолжаются в виде выступлений заинтересованных участников на публичных площадках СМИ и публикаций в научных журналах. В статье в журнале «Образовательная политика» депутат Государственной Думы РФ О.Н. Смолин показывает, что парламентский контроль в части школьного образования достаточно слаб, проблемы решаются в течение долгого времени: в течение года из двенадцати рекомендаций ГД РФ от 23.07.2019 № 6683-7-ГД выполнена полностью только одна, а частично – две [3].

Значимыми вовлеченными акторами в процесс конструирования социальных проблем становятся общероссийские общественные организации и движения, а также региональные структуры гражданского общества.

Общественная палата Российской Федерации в ежегодных докладах о состоянии гражданского общества представляет состояние и перспективы общественных движений, в том числе на селе, действующие в рамках развития сельских школ.

В 2015 г. в докладе отмечено существенное отставание сельских территорий от городских; по количеству успешно функционирующих НКО и тем более получающих финансовую поддержку Фонда Президентских грантов [4]. По мнению авторов, только сильно функционирующий третий сектор способен вовлечь население территории в совместное решение проблем, определение нужной инфраструктуры и получение государственных субсидий на ее поддержание (в том числе сельских школ). В докладе 2016 г. представлены результаты работы Общественной палаты РФ по консолидации ресурсов поддержки предпринимательства на сельских территориях «Карта возможностей». Вместе с многими позитивными итогами приведены данные опроса активистов, которые подтверждают стабильное функционирование школ в части сельских поселений [5]. Доклад 2017 г. подчеркивает необходимость развития мелких и средних фермерских хозяйств, которые лучше крупных агрохолдингов заботятся именно о сохранении села и развитии населения. В качестве примера приводится случай сохранения сельской школы от «оптимизации» на хуторе Восточный в Краснодарском крае [6]. В 2018 г. текст посвящен качеству подготовки педагогов, но не выделяет трудности закрепления специалистов в сельских школах. Основное внимание в отношении села уделено качеству жизни в сельской глубинке и разработке Стандарта села с требованиями к администрации по направлениям «Доступность объектов социальной сферы», «Обеспеченность жильем и его состояние» [7]. Анализ проблем в докладе 2019 г. обращен к поддержке инициатив заниматься несельскохозяйственными видами деятельности селян, а вопросы развития школьного образования увязаны с совершенствованием федеральных государственных стандартов [8]. А вот уже в 2020 г. из-за пандемии тема школы возвращается в обсуждение только в связи с дистанционными технологиями [9]. Таким образом, внимание к процессам, происходящим на селе, возобновляется практически каждый год, в том числе благодаря обращениям граждан.

Общественное движение «Общероссийский народный фронт» разворачивает проект «Школа большой страны» [10] – сообщество для директоров школ, благодаря которому они могли бы обмениваться успешным управленческим опытом, а вместе с Российским союзом сельской молодежи – проект по развитию системы агроклассов в сельских школах [11]. ОНФ отвечает на запросы от конкретных школ (ремонт туалета в сельской школе Брянской области) [12].

Именно такие площадки способны привлечь внимание представителей федеральных и региональных профильных министерств к решению вопросов поддержки педагогических работников на селе, хотя пока в круг проблем ОНФ включены отдельные вопросы реформирования сельских школ.

Наиболее адресным по отношению к закреплению педагогических работников на селе является масштабный общественный проект «Сельские школы России». В настоящее время портал, организованный как собрание сельских школ (энциклопедия, свод знаний об активности сельских школ), стал действующей площадкой для формирования развивающих программ, реализации грантовых проектов, мер помощи и поддержки конкретных сельских школ и взаимодействия с партнерами различных уровней. В рамках развития сообщества портал объявил об открытии трибуны «Лидеры мнений по развитию сельской школы» для публикации выступлений руководителей сельских школ, которые ведут большую просветительскую деятельность. Пока в этой рубрике представлены всего два материала, но впоследствии она может развернуться в открытое заинтересованное обсуждение самых насущных проблем сельских педагогов.

Общероссийская молодежная общественная организация «Российский союз сельской молодежи» насчитывает уже более 12 лет существования. В настоящее время это единственная организация, представляющая интересы сельской молодежи восьмидесяти субъектов РФ на всероссийском уровне.

РССМ входит в центральный штаб ОНФ, Общественную палату Российской Федерации; общественный совет при Минсельхозе России; рабочую группу Государственного совета Российской Федерации «Молодёжная политика»; федеральный реестр молодёжных и детских объединений Росмолодёжи; реестр некоммерческих организаций – исполнителей общественно полезных услуг Минюста России. Основные значимые проекты РССМ в сотрудничестве с ОНФ, министерством сельского хозяйства РФ – «Инкубатор сельских инициатив», «Село. Территория развития».

РССМ выстраивает социальное партнёрство и международные отношения: проведены два международных форума сельской молодёжи, реализован проект «Взаимодействие сельской молодёжи России и стран Европы», подписано соглашение о сотрудничестве с Немецким союзом сельской молодёжи, Белорусским республиканским союзом молодёжи, Ассоциацией молодых фермеров Голландии. Совсем недавно были подписаны соглашения с молодёжными организациями Казахстана и Киргизии. Активисты союза являются постоянными участниками международных мероприятий и встреч с иностранными коллегами.

Союзом ведётся активная проектная деятельность: за период с 2008 по 2019 г. реализовано более 100 социально значимых проектов, в которых приняли участие свыше 5 миллионов человек [13].

В представленной работе авторами был проведен анализ постов учителей – участников РССМ в социальной сети «ВКонтакте» (открытые данные) за период с 01 сентября по 31 декабря 2020 г. (отобрано более 600 текстов из 2 000 первоначально собранных).

В результате исследований были выбраны не более 50 постов, которые содержательно можно отнести к индивидуальной активности в пользу развития села или сельских специалистов. Остальные тексты – официальные объявления, пересланные в своей ленте, ситуативные посты о рецептах и участии в конкурсе, комментарии фотографий на отдыхе.

Мы с ребятами из РССМ запустили марафон для студентов-аграриев. Это отличный шанс проявить себя! РССМ отберет лучшие видеовизитки и уже 9 ноября начнется этап

онлайн-голосования во ВКонтакте. Победители получают рекомендацию по стажировке в компанию мечты, дипломы, призы с символикой от РССМ!

#МаминаИсторияУспеха2020 Наша семья участвует в данном проекте! Всероссийские конкурсы «История успеха». Подробная информация на сайте <https://многодетные-семьи.рф/тата>

Внимание!!! Совсем скоро!!! В эту субботу!!! 29 августа в 16:00 ч. Состоится День села Трубино "Украсим Трубино цветами"!!!!

Как правило, официальные объявления переходят из поста в пост, не слишком часто встречаются тексты, выражающие личную позицию автора – среди 50 отобранных большинство (38) – об участии в конкурсе своей семьи, школы, учеников.

Первый анализ постов позволяет сказать о том, что молодые подписчики паблика РССМ из числа сельских педагогических работников практически не упоминают в этом пространстве собственных социальных проблем и проблем развития села или школы. Сообщество в социальной сети не воспринимается авторами как публичная площадка, на которой могут встретиться заинтересованные стороны или создаться группа по интересам, которая впоследствии будет выдвигать утверждения-требования по поводу закрепления молодых специалистов сферы образования на селе.

Представляется важным найти для педагогических работников на селе дополнительные возможности карьерного роста, внеучебного развития, продвижения по линии социального активизма. Всероссийское детско-юношеское военно-патриотическое общественное движение создано в 2016 г. по инициативе Министра обороны РФ Сергея Шойгу. В настоящее время в рядах Юнармии уже состоит 775 тысяч школьников [14], к 9 мая 2021 г. прогнозировалось и достигнуто количество 1 млн человек.

В Юнармии довольно массово представлены именно сельские школы, и работа ячеек на селе предоставляет учителям и администрации сельских школ некоторые бонусы от региональных министерств образования и в рамках проектов Юнармии: слеты руководителей ячеек и отрядов, конкурсные мероприятия с выездом в детские организации отдыха, обучение, повышение квалификации. По мнению директоров сельских школ, если в составе коллектива есть активный руководитель отряда Юнармии, «жизнь вокруг такого бьет ключом».

Нижегородское региональное отделение Юнармии зарегистрировано в 2016 г. и сейчас включает 61 местное отделение – по количеству муниципалитетов в области. Состав управления – 1 начальник регионального отделения и 61 начальник местного отделения (в каждом территориальном образовании), а также 610 руководителей отрядов (это сотрудники – учителя, которые непосредственно занимаются с детьми).

С 2016 г. были зарегистрированы 61 местное отделение, но работают в течение пяти лет только 20 отделений, и каждый год в активную деятельность вступали отряды муниципалитетов: в 2016 г. – 20, в 2017 г. – 30, в 2018 г. – 11 отделений.

Сейчас в Юнармии Нижегородской области 23 000 юнармейцев; динамика прироста численности этой организации такова: 2016 г. – 1 000 человек, 2017 г. – 3 000, 2018 г. – 7 000, 2019 г. – 21 000, 2020 г. – 22 000, 2021 г. – 23 000 человек (на начало года) [14].

В передовиках числятся Казань (28 000 юнармейцев) и Ульяновск (29 000 юнармейцев); Москва опережает всех – 30 000 участников.

Гражданское общество на сельских территориях России пока развито достаточно слабо, и сельская школа не так часто находит для себя поддержку в структурах третьего сектора на селе. Тем важнее для конструирования социальной проблемы закрепления специалистов сферы образования на селе общероссийские публичные арены, предоставляющие возможность продвижения утверждений-требований, объединяющих максимальное количество заинтересованных сторон и значимых для ведомственных структур.

Парламентские слушания по вопросам развития сельских школ стали наиболее ресурсной публичной ареной и максимально представительной по числу участвующих сторон (от учителя – до представителя Совета Федерации РФ).

Общественные структуры всероссийского уровня предоставляют возможность донести утверждение-требование сельских педагогов до самого высокого уровня власти (ОП РФ, ОНФ РФ). Общественные союзы и движения ориентируют сельских педагогов объединиться для обмена опытом разрешения проблем конкретных школ и тиражирования лучших практик, а также предоставляют им дополнительные ресурсы для развития и вертикальной мобильности (РССМ, Юнармия, РДШ, волонтерское движение) как в иерархии систем образования, так и внутри этих общественных структур.

На фоне продолжающегося реформирования сельских школ и недостаточной наполненности муниципальных бюджетов структуры гражданского общества на протяжении последних пяти лет очень активно включаются в конструирование проблем сельской школы и могут сыграть решающую роль в закреплении педагогических работников на селе.

Список источников

1. Федоров Н.В. Стенограмма парламентских слушаний на тему «О состоянии и перспективах развития сельской школы в Российской Федерации». URL: [council.gov.ru>media/files/...pdf](http://council.gov.ru/media/files/...pdf)
2. Постановление Государственной Думы Федерального собрания Российской Федерации от 23.07.2019 № 6683-7 ГД «О рекомендациях парламентских слушаний на тему «О мерах по повышению качества образования в Российской Федерации». URL: [docs.cntd.ru>document/560764533...pdf](http://docs.cntd.ru/document/560764533...pdf)
3. Смолин О.Н. Слушали, рекомендовали, забыли. Долгий ящик системных решений // Научно-публицистический журнал «Образовательная политика». URL: <https://edpolicy.ru/system-solutions.pdf>
4. Доклад о состоянии гражданского общества в Российской Федерации за 2015 год. М. : Общественная палата Российской Федерации, 2015. С. 27–34. URL: pdf.knigi-x.ru/...doklad...sostoyanii-grazhdanskogo...v...pdf
5. Доклад о состоянии гражданского общества в Российской Федерации за 2016 год. М. : Общественная палата Российской Федерации, 2016. С. 42–58. URL: lib.knigi-x.ru/...doklad...sostoyanii-grazhdanskogo...v...pdf
6. Доклад о состоянии гражданского общества в Российской Федерации за 2017 год. М. : Общественная палата Российской Федерации, 2017. С. 77. URL: report2017.oprf.ru/pdf
7. Доклад о состоянии гражданского общества в Российской Федерации за 2018 год. М. : Общественная палата Российской Федерации, 2018. С. 23. URL: report2018.oprf.ru/pdf
8. Доклад о состоянии гражданского общества в Российской Федерации за 2019 год. М. : Общественная палата Российской Федерации, 2019. С. 114. URL: report2019.oprf.ru/pdf
9. Доклад о состоянии гражданского общества в Российской Федерации за 2020 год. М. : Общественная палата Российской Федерации, 2020. С. 100. URL: report2020.oprf.ru/pdf
10. Проект ОНФ: Школы большой страны. URL: [onf.ru>project/90445/news/pdf](http://onf.ru/project/90445/news/pdf)
11. Эксперты ОНФ и РССМ сформируют «дорожную карту» развития системы агроклассов в сельских школах. URL: <https://onf.ru/2020/08/21/eksperty-onf-i-rssm-sformiruyut-dorozhnyuyu-kartu-razvitiya-sistemy-agroklassov-v-selskih/>
12. Новости ОНФ. URL: <https://onf.ru> > 2021/01/12 > posle-signal-a-bryansk-...>
13. Официальный сайт РССМ. URL: <https://rssm.su/company/>
14. Официальный сайт ВВПОД «ЮНАРМИЯ». URL: <https://yunarmy.ru/headquarters/about/>

RURAL SCHOOL: AN ANALYSIS OF PROBLEM PRESENTATION IN PUBLIC ARENAS

E.S. Belyaeva, I.E. Petrova

*Lobachevsky State University of Nizhny Novgorod – National Research University,
Nizhny Novgorod, Russian Federation*

Abstract. The topic of preserving and developing rural schools in modern Russia is a hot one, but problems of rural teachers and the importance of the school for rural society are presented mainly within the work of state authorities and institutions of departmental subordination to the sphere of education. Participation of teachers and villagers (schoolchildren’s parents and activists) in this process is difficult due to limited access. The aim of the study is to find out the actors and ways of constructing opinions about the problems of rural schools in various public arenas, including social networks (pages of communities dedicated to supporting rural initiatives). Conclusions are made about the lack of rural educators’ activity in the promotion of the position of the school community about the resources and barriers to the retention of educational personnel in rural areas.

Keywords: rural schools, teachers, rural society, problem construction, public arenas, social networks

ИЗ ОПЫТА ОРГАНИЗАЦИИ АВТОМАТИЗИРОВАННОГО СБОРА ДАННЫХ В ТОМСКОМ УНИВЕРСИТЕТЕ

А.Л. Благинин, Э.Р. Сайфулин, А.Ю. Саркисова

Национальный исследовательский Томский государственный университет, Томск, Россия

Аннотация. В статье систематизированы методы, инструменты, источники, примеры результатов автоматизированного сбора данных больших объёмов, актуализированные в практике проектов по социально значимым тематикам Центра прикладного анализа больших данных НИ ТГУ в 2016–2022 гг. Охарактеризованы выгрузка данных с помощью API, веб-скрейпинг, эмуляция работы мобильного приложения, использование специализированных платформ. Отмечено значение инструментов машинного обучения для анализа данных на этапе выгрузки. Названы возможности систем управления базами данных для организации работы с выгруженными данными. Обозначены проблемы и ограничения автоматизированного сбора данных. Приведены некоторые примеры источников данных. Особое внимание уделено социальной сети «ВКонтакте» как источнику, для которого на текущий момент в Центре имеется наиболее разработанный инструментарий. Приводятся примеры описаний наборов данных, а также примеры методики сбора. Статья адресована специалистам в области социальных и гуманитарных наук. Цели статьи связаны с популяризацией актуальных способов и ресурсов работы с большими данными и расширением их использования в исследовательской практике.

Ключевые слова: автоматизированный сбор данных, большие данные, коллекция данных, набор данных, веб-скрейпинг, открытый API, социальные сети, цифровая гуманитаристика

Введение

Традиционные методы сбора данных (интервью, опросы, анкетирование, прототипирование и др.) остаются востребованными в разных областях – от социологии [1] до проектного менеджмента [2].

Тем не менее форсированная цифровизация современной жизни и увеличивающиеся потоки информации закономерно привели к той роли больших данных в перспективе развития научных исследований, которую они начинают занимать в соответствии как с декларируемыми в официальном дискурсе приоритетами стратегического развития, так и с реальными задачами самых разных академических и прикладных областей научного знания. Огромное количество данных генерируется каждый день на различных сайтах, коммуникативных площадках сети Интернет, которая сегодня является самым массовым источником оперативной информации. Особенность её – то, что рядовой пользователь часто выступает создателем и получателем информации одновременно.

При работе с большими данными ручные сбор и обработка становятся невозможными, что требует от научного сообщества освоения новых, автоматизированных, методов и инструментов.

Исследование больших данных представляет собой многоступенчатый процесс, этапы которого описаны, например, в [3]. Одним из первых этапов (предшествуют ему постановка цели и задач, планирование) является этап сбора данных.

Важность этапа сбора, обеспечивающего доступ к данным и определяющего полноту и релевантность выборки, обуславливает устойчивый интерес к нему в научной литературе. Исследователи поднимают вопросы разработки универсальных инструментов для извлечения информации с разнотипных сайтов, стратегии обхода защиты сайтов, технические средства фильтрации ботов, апробации алгоритмов и тестирования программного обеспечения для автоматического сбора информации и др. В качестве источников данных используются, например, веб-сайты СМИ [4], электронные архивы журналов [5], популярный сайт объявлений Avito [6], крупнейшая в мире библиографическая база данных WorldCat [7] и др. Ряд

статей посвящён платформам по сбору (и анализу) данных: например, [8]. Затрагиваются также технические вопросы корреляции сбора и анализа разных этапов работы с данными: например, Tikito с соавторами показывают, что использование интеллектуального процесса сбора данных в мобильных устройствах может обеспечить экономию с точки зрения хранения и анализа данных на облачном уровне [9].

В настоящей статье систематизируется опыт Центра прикладного анализа больших данных НИ ТГУ (далее – Центр) в области автоматизированного сбора данных больших объёмов. Характеризуются основные способы извлечения данных, проблемы и ограничения при автоматизированном сборе, источники данных, оборудование, приводятся примеры полученных наборов данных и ссылки на исследования на их основе.

Внимание фокусируется на типе данных, ориентированных на социальные науки. Основным источником таких данных является цифровой след человека в сети Интернет.

В некоторых научных областях достаточно основательно разработана методология сбора, систематизации и организации данных в целом. Так, в доказательной медицине на данных организуются систематические обзоры с количественным обобщением, метаанализы, обеспечивающие синтез научной медицинской информации для принятия необходимых клинических решений (см., например, [10]). Установка на конвенциональные методики и унификации методов систематического анализа приводит к появлению соответствующих государственных и ведомственных стандартов.

В науках об обществе, хотя ресурс больших данных в них очень востребован [11], методология сбора и обработки разработана существенно слабее. Более дискуссионными оказываются степень достижения однородности, полноты, достоверности данных; чаще возникают вопросы, связанные с концептуальным единством выборки.

В практике Центра наиболее освоенным источником данных на настоящий момент являются социальные сети. Преимущества источника – популярность и массовость, как правило, открытость и доступность, добровольность и непринуждённый (естественный) характер коммуникации, неограниченный круг тем. Ряд аспектов автоматизированного сбора данных рассматривается в статье на примере социальной сети «ВКонтакте».

Цели статьи связаны с популяризацией сведений о методах и ресурсах работы с большими данными в перспективе расширения их использования в социальных и гуманитарных исследованиях.

1. Методы и инструменты автоматизированного сбора данных

В практике автоматизированного сбора данных с интернет-страниц используется два основных метода: выгрузка посредством API и веб-скрейпинг. Дополнительно могут быть отмечены некоторые другие способы обеспечения доступа к данным.

1. Удобным и экономичным признаётся **сбор данных с помощью API** (Application Programming Interface) – прикладного интерфейса программирования. При использовании API осуществляется выгрузка данных через предоставленные для этого документированные методы на самом сайте.

Техническая проблема возникает в связи с тем, что API есть далеко не у всех веб-сервисов. Если разработчик/владелец сайта заинтересован в том, чтобы с контентом сайта работали автоматически (например, в интеграции с приложениями), он разрабатывает API и предоставляет доступ пользователям. Таким образом он одновременно делает информацию сайта доступной для аналитики посредством её автоматизированного сбора.

Обычно различают открытый и закрытый API. Закрытый API создан для внутренней разработки и поддержания продукта, в то время как публичный, или открытый, API позволяет пользователям интегрировать свои собственные приложения с продуктом.

В качестве примера проекта Центра, основанного на данных, полученных с помощью API, можно привести расчёт индекса субъективного качества жизни для 83 регионов России-

ской Федерации на основе 19 параметров экономических, социальных и политических аспектов качества жизни, отражённых в постах пользователей социальной сети «ВКонтакте» [12].

2. **Веб-скрейпинг** – способ выгрузки данных, который представляет собой имитацию работы человека в браузере. Скрейпер – это программа, позволяющая извлекать содержимое сайта посредством обработки его веб-страниц. Такие действия, как переход по ссылкам, сканирование текста и т.п., осуществляются автоматически.

Веб-скрейпинг является достаточно ресурсоёмкой технической процедурой, требует навыков web-программирования. Задачи специалиста включают, в частности, обработку исходного текста страницы; обход блокировок; работу с интерактивными элементами страницы (эмуляцию действий пользователя: каждому клику, слайдеру и т.п. соответствует свой фрагмент кода). Обязательным является знание языка программирования Java Script. Так как все сайты разные с точки зрения структуры и дизайна, под каждый сайт нужно писать свой скрейпер. Если сайт меняет логику своей работы (например, интерфейс), возникнет необходимость писать новую программу.

Однако недостатки веб-скрейпинга компенсируются тем, что он применим абсолютно ко всем сайтам. Программист, осуществляющий веб-скрейпинг, открывает просмотр кода страницы (язык html) (при этом он иногда вынужден от самого корневого фрагмента кода искать необходимый текст): закрыть доступ к данной странице технически невозможно, так как если нельзя просматривать страницу с кодом, то нельзя увидеть и сам сайт.

Итак, и для применения API, и для веб-скрейпинга нужны навыки программирования, но использование API – технически более простой метод (для него тоже пишется код, но он значительно короче), освоить его не IT-специалисту проще. Веб-скрейпинг однозначно требует специализированных навыков. С другой стороны, не у всех сайтов есть API и не у всех есть API, позволяющие выгружать данные (например, у новостных сайтов чаще всего нет API), в то время как веб-скрейпингу может быть подвержен любой сайт (при наличии достаточных компетенций).

Примером проекта Центра, в котором данные извлекались с помощью веб-скрейпинга, может служить исследование концепта «политик» в современных китайских газетах на английском языке: для него из 6 крупнейших газет были извлечены все новостные тексты за 2020 г., включающие слово «politician» и/или его синонимы.

3. В качестве отдельного метода сбора данных может быть выделена **эмуляция работы мобильного приложения**. Сбор данных с сайтов, при взаимодействии с которыми нельзя использовать API и особо трудозатратно осуществлять веб-скрейпинг, может выполняться с помощью программы, имитирующей работу мобильного устройства.

Например, извлечь данные Instagram (аналогично – Twitter, TikTok и др.) непосредственно через веб-сайт максимально сложно из-за жёсткого противодействия скрейпингу. Но возможно сделать это через эмулятор – мобильный ОС Android и установленное в нём целевое приложение.

Данный способ технически ещё сложнее, чем веб-скрейпинг, он также является более дорогим в связи со спецификой мобильной разработки. Но метод является перспективным и будет интенсивно развиваться в практике сбора данных.

4. Доступ к данным для широкого круга пользователей может обеспечивать **использование специализированных платформ**.

Платформы, в данном случае представляющие собой среды выполнения определённых алгоритмов сбора данных, позволяют упростить автоматизированный поиск и выгрузку данных и сделать их доступными для специалистов социогуманитарного профиля, не владеющих навыками программирования.

В практике Центра используются следующие платформы.

✓ Платформа сбора и обработки данных социальных сетей SNLab (<https://lk.opendata.university/>) [13]. Платформа позволяет выгружать все типы данных социальной сети (о типах данных см. далее) в режиме реального времени. Выгрузка осуществляется либо по заданным ключевым словам, либо по прямой ссылке.

Платформа позволяет решать, например, такие задачи: выгрузить все томские сообщества с заданным порогом по количеству подписчиков; или выгрузить профили всех подписчиков заданного списка сообществ, отобранных по определенному тематическому критерию.

Примером исследования Центра, экспериментальный материал для которого был получен с помощью данной платформы, является сетевой анализ структуры деструктивных девиантных сообществ, транслирующих субкультуру скулшутеров [14].

✓ Аналитическая платформа PolyAnalyst (<https://www.megaputer.com/ru/polyanalyst/>) (разработчик – российская компания «Megaputer Intelligence», индустриальный партнёр НИ ТГУ) является востребованным инструментом текстовой аналитики. В первую очередь, она предназначена для обработки и анализа текстовых данных, но поддерживает также функции автоматизированного сбора. Платформа позволяет, например, парсить интернет-страницы или выгружать данные из социальных сетей.

✓ Программно-аппаратный комплекс «Крибрум» (<https://www.kribrum.ru/>) (разработчик – АО «Крибрум», индустриальный партнёр НИ ТГУ) позволяет осуществлять мониторинг и сбор данных русскоязычных средств массовой коммуникации. Пользователь создаёт поисковый объект, инструмент собирает все упоминания СМИ о данном объекте, начиная с момента создания, с учётом метаданных (источник, автор, дата и др.).

Используются также более специализированные авторские платформы.

✓ Платформа психологической диагностики с авторизацией респондентов в социальной сети «ВКонтакте» (<https://ivik.org/>) [15] позволяет осуществлять сбор информации о психологических характеристиках идентифицированных пользователей социальной сети «ВКонтакте»: респонденты проходят опросы и привязывают их к своему аккаунту «ВКонтакте».

С помощью данной платформы, например, проанализирована связь психологических характеристик школьников и студентов – пользователей социальной сети «ВКонтакте» – с подписками на сообщества с девиантным контентом и временем активности в социальной сети [16].

✓ Платформа для рекрутинга абитуриентов (<http://talent-search.opendata.university/login>) разрабатывалась для приёмных комиссий университетов (для поиска целевой аудитории в социальной сети). Платформа позволяет осуществлять гибкий поиск по уже подготовленным данным. Выгрузка ведётся по таким параметрам, как возраст, год выпуска, регион, пол, когнитивные характеристики, образовательные подписки, образовательные интересы.

5) Исследователям не стоит забывать про самый простой и очевидный способ получения доступа к данным – **предоставление их непосредственно владельцем Интернет-ресурса или организацией – оператором данных**. Так, например, все крупные и давно существующие вузы одновременно являются крупными операторами данных (речь не идёт о персональных данных). Данные также могут предоставлять заказчики прикладных исследований.

Владелец данных / уполномоченное лицо даёт разрешение, разработчик (со стороны владельца) обеспечивает доступ к данным: реквизиты для входа на сервер, либо непосредственное предоставление данных в требуемом формате. Данные могут быть предоставлены в широком наборе форматов; одними из самых востребованных являются форматы XLSX, CSV.

Необходимо отметить, что при автоматизированном сборе данных могут быть задействованы **инструменты машинного обучения**, которые дают возможность анализировать содержимое на этапе выгрузки.

Такой подход позволяет проводить фильтрацию данных на самом раннем этапе и работать с трудно поддающимися анализу типами данных, в частности, это касается разбора текстов и лингвистического анализа.

Например, стоит задача выгрузить из социальной сети «ВКонтакте» все тексты в жанре автобиографии. Для этой цели на заранее подготовленном наборе данных, содержащем раз-

личные виды текстов и характеристику их жанра, производится обучение модели. Когда достигнут необходимый уровень точности предсказания модели, можно запустить процесс выгрузки, в ходе которого обученная модель будет давать свою характеристику подаваемым в неё текстам.

Хранение собранных данных и последующий доступ к ним организуется при помощи **систем управления базами данных (СУБД)**. При подготовке базы данных разработчиками разворачивается сервер СУБД, создаётся база данных, которая автоматически наполняется. База данных отличается от датасета тем, что обычно она управляется посредством программного обеспечения (СУБД).

СУБД позволяет создавать выборки по ключевым словам, удалять, изменять часть данных, а также выполнять сложные аналитические операции. СУБД включает в себя функцию очистки данных (можно задавать условия фильтра, например, извлечение текстов объёмом не менее 30 символов).

В настоящее время Центр имеет полную базу данных социальной сети «ВКонтакте», на её основе оформлено несколько результатов интеллектуальной деятельности [17–19].

Для решения конкретных исследовательских задач возможно извлечение материала из имеющейся базы данных. Например, для извлечения постов пользователей о Победе в Великой Отечественной войне в период празднования 75-летия Победы была использована уже скачанная база данных «ВКонтакте»: был запущен поиск тематических сообщений с опорой на список лингвистических маркеров (ключевых слов). В период с 01.01.2019 по 17.08.2020 обнаружено и извлечено 567 623 поста, тематически связанных с Победой.

2. Проблемы, ограничения, условия автоматизированного сбора данных

Обозначим основные ограничения, стоящие перед научными коллективами, которые реализуют проекты, требующие сбора больших массивов данных.

1. Автоматизированный сбор данных требует специального оборудования. В зависимости от объема выгружаемых данных требуется сервер с достаточным объёмом дискового пространства и вычислительной мощностью, чтобы извлеченные данные хранить и обрабатывать. В практике работы Центра используются суперкомпьютер СКИФ Cyberia и ряд менее мощных серверов для сбора и обработки данных.

2. Другим условием, уже упоминавшимся выше, является необходимость специальных навыков. IT-специалисты плохо задерживаются в научно-образовательных организациях, и это серьёзная проблема.

3. Технические ограничения и трудности во многом связаны с защитой сайтов от автоматизированного сбора данных. Автоматизированный сбор данных сопряжен с большим числом запросов в единицу времени, поэтому причиной установления сайтами защиты от веб-скрейпинга не обязательно является защита собственно данных: часто сайты стремятся сократить количество запросов и снизить нагрузку на сайт.

В большинстве случаев API нельзя использовать без авторизации, то есть его возможно использовать только от имени пользователя. Для этого создаются многочисленные боты – функционирующие на коммуникативных площадках программы, имитирующие поведение человека.

Самые популярные сайты защищаются от веб-скрейпинга, разрабатывая методы определения автоматических запросов (например, учёт частоты запросов в секунду, активность мышки и др.).

В случае распознавания скрейпера может быть заблокирован IP-адрес. Но, как правило, адрес блокируется временно с целью лишь замедления работы скрейпера. Заблокировать адрес на постоянной основе обычно невозможно, так как с одного IP-адреса работают и добросовестные пользователи, и скрейперы.

В целом, несмотря на совершенствование методов защиты, «в настоящий момент невозможно полностью предотвратить автоматизированный интеллектуальный сбор данных с

ресурсов сети Интернет» [6, с. 83]. Правовые и этические аспекты сбора данных на сегодняшний день остаются дискуссионными.

Автоматизированному сбору подлежат только публичные данные (например, только открытые профили и страницы социальной сети), то есть те, которые можно извлечь и «вручную».

С закрытыми профилями пользователей и закрытыми страницами сообществ связана и содержательная проблема: соответствующие данные не могут быть учтены в общей статистике, а также при социально-сетевом анализе, где большое значение для достоверности результатов имеет полнота выборки. Например, если в конкретном сообществе с открытыми постами скрыты подписчики, то они не будут отражены в исследованиях с построением сетевого взаимодействия сообществ за счёт общих подписчиков. В графе сообществ такое сообщество автоматически попадает в (ложные) изоляты.

При работе с персональными данными все данные обезличиваются.

Сайты с открытыми API заключают пользовательские соглашения с субъектами данных.

В ряде случаев статьи с результатами на основе больших данных проходят этический комитет перед публикацией.

3. Источники данных. Социальная сеть «ВКонтакте» как источник данных. Типы данных, извлекаемых из социальной сети

В качестве источников данных могут выступать любые ресурсы сети Интернет, открытые субъектом данных для просмотра.

Социальные медиа (в том числе социальные сети и сетевые СМИ) используются в практике социогуманитарных исследований Центра наиболее часто [12, 14, 20–22].

Многоязычная Интернет-энциклопедия «Википедия» служит источником в продолжающемся проекте о расчёте степени влияния университетов на общество [23].

Имеется также опыт работы со следующими источниками: международные базы научных публикаций; сайт Интернет-рекрутинга HeadHunter (<https://hh.ru/>); сайт Единой информационной системы в сфере закупок (<https://zakupki.gov.ru/>), база резюме и вакансий для студентов ТГУ UniProfi; внутренние данные ТГУ (система Moodle, портал РНД).

До текущего времени в прикладных и научных проектах Центра наиболее использующимся и освоенным источником данных выступала **социальная сеть «ВКонтакте»** (<https://vk.com/>). Причиной ее популярности стало повышение роли социальных сетей в современной коммуникации, распространении информации и формировании общественного мнения (97 миллионов посетителей ежемесячно; открытый API, отличающий данный ресурс от ряда других социальных сетей).

API позволяет специалисту по сбору данных извлекать данные в исходном виде. Впоследствии он передаёт их заказчику в требуемом формате для последующего анализа. В подавляющем большинстве случаев данные передаются в виде таблицы, где одна строка представляет один экземпляр объекта, а колонки таблицы – его атрибуты.

В таком источнике данных, как «ВКонтакте», есть три основных объекта данных: пользователь, сообщество и пост. Информация об этих объектах представляет собой данные, которые могут автоматически извлекаться, обрабатываться и затем анализироваться, давая информацию о массовых характеристиках современного цифрового общества.

Данными группы «Пользователь» являются, во-первых, данные, введённые самим пользователем, например: имя, отчество, фамилия, дата рождения, семейное положение, город, место учёбы и / или работы (включая предыдущие), персональный сайт, родной язык, родственники, образование (включая годы обучения, учебные заведения, факультет, специальность, форму обучения), политические предпочтения, декларируемые ценности и особенности мировоззрения, отношение к курению и алкоголю, интересы, предпочтения в музыке, фотографии, видео. Пользователь может указать ложные данные (например, пользователи

часто заведомо неверно указывают свой город) – неподлинный характер данных в ряде случаев может быть установлен, но требует дополнительных (иногда весьма ресурсоёмких) операций. В то же время особенности самопозиционирования пользователя, системные закономерности вымышленных данных в больших масштабах могут выступать предметом отдельного интереса. Во-вторых, к данным о пользователе относятся более объективные сведения: его ID, друзья, подписки (группы и сообщества).

При автоматизированном сборе возможны отбор / классификация пользователей по каждому из перечисленных критериев, а также по их комбинациям. Например, при помощи агрегации возможно установить количество пользователей, проживающих в каждом регионе, для каждого пола. При этом список атрибутов для агрегации может быть расширен. На основе этих данных дата-аналитики получают интересные результаты о социально значимых, психологических и прочих закономерностях цифрового общества.

Данные группы «Пользователь» выгружаются при условии открытого профиля. Если пользователь закрыл профиль, данные в полной мере быть извлечены не могут.

Сообщество – объект социальной сети «ВКонтакте» – группа или публичная страница, где активно обсуждаются различные аспекты жизни и актуальные новости разных тематик. **Данными группы «Сообщество»** являются ID сообщества, его название, описание, доменное имя, количество подписчиков, тип сообщества (event, public, group: первый из них привязан к определенной дате и посвящён какому-либо событию – например, акции, организованному мероприятию и др.; второй и третий различаются возможностью / невозможностью любому участнику сообщества публиковать посты под своим именем, а не от имени сообщества; в публичных страницах (public) более заметную роль играет модерация, цензура), местоположение, статус, тематика сообщества (средствами API может быть извлечена с полем activity; для «событий» – содержит дату их проведения).

Владелец сообщества тоже может скрыть стену сообщества и / или своих подписчиков. Последнее, как замечалось выше, делает невозможным, в частности, учет данного сообщества в сетевом анализе взаимодействия сообществ (построении сети общих подписчиков разных сообществ).

В социальной сети «ВКонтакте» на 2022 г. имеется более 200 миллионов сообществ, из них более 8 миллионов имеют не менее 100 подписчиков.

Данными группы «Пост» являются само сообщение (текст, изображение, аудио или видео), ID поста, ID автора поста, ID владельца сообщества, ID предложившего пост лица (для сообщества), количественные показатели поста: количество просмотров, лайков, репостов, комментариев. Интересный тип данных представляет собой «сору-history»: история репостов, сопровождаемых многоступенчатыми дополнениями от каждого лица, использовавшего функцию репоста. К группе данных о посте также примыкают комментарии, каждый из которых, в свою очередь, сопровождается аналогичным посту набором метаданных.

Для дата-аналитика особое значение имеет ID (пользователя, сообщества или поста), так как его нельзя ни подделать, ни скрыть. ID – это уникальные числовые идентификаторы, которые имеют сплошную нумерацию, начиная с единицы, то есть сохраняют последовательность возникновения нового пользователя сети, сообщества или поста в реальном времени.

Профиль пользователя, сообщества или пост могут быть удалены их владельцем, в таком случае его увидеть и извлечь специалист по сбору данных не сможет (если он не был им скачан ранее). Однако при этом ID удалённого объекта сохраняется; таким образом, видно, сколько объектов было удалено.

4. О методиках сбора данных

Итак, современные технологии позволяют извлекать информацию о пользователях и, например, создавать коллективный социально-демографический портрет представителя цифрового общества, соответствующего заданным критериям.

Извлечённые данные о сообществах позволяют выявлять основные коммуникативные площадки в социальной сети по определённой теме, строить сети взаимосвязи сообществ.

Извлечённые посты представляют собой материал непосредственно для анализа контента социальной сети.

Наиболее частой оказывается задача извлечения данных по тематическому критерию.

Основным инструментом при этом служит полнотекстовый поиск по лингвистическим маркерам, то есть ключевым словам. Исследователь разрабатывает список ключевых слов – результатом автоматизированного сбора станут посты, включающие хотя бы одно из этих слов (или, при необходимости, их комбинации), независимо от словоформы. Возможность поиска различных форм слова достигается за счет стемминга и лемматизации – приведения слов к их исходной форме и отсечения окончаний.

При сборе данных по ключевым словам может быть использован метод снежного кома: из полученной выборки по составленному пользователем списку слов далее уже автоматически извлекаются ключевые слова на основе метрик частотности и значимости, далее процедура повторяется на вновь полученной выборке и т.д. Метод позволяет получать качественные выборки с высокой степенью полноты. При разработке лингвистических маркеров можно также сделать список стоп-слов: если текст включает слово из списка, он в выборку не попадает. Стоп-слова выполняют роль фильтра, позволяя предупредить в какой-то мере последующий этап очистки данных.

При подготовке технического задания на сбор данных важным моментом является критерий ограничения материала.

В практике реальных исследований работа с социальной сетью целиком не производится ввиду нецелесообразности и чрезмерной трудоёмкости обработки такого объёма информации. Эффективным является подход, основанный на переходе от общих критериев поиска материала к более частным – на выборках данных ограниченного объёма, которые после предварительного анализа можно расширять.

Так, могут быть использованы ограничения по региону, периоду; поиск по ключевым словам может производиться не по постам, а по названиям и описаниям сообществ.

Для ряда задач окажется приемлемым построение случайной выборки, например, когда требуется описать генеральную совокупность с заданной точностью. Случайные выборки создаются классическим для прикладной статистики способом – сэмплированием.

Ниже приведены примеры наборов данных по некоторым продолжающимся проектам Центра с описанием методики их получения. В двух из них источником выступает социальная сеть «ВКонтакте», в одном – интернет-энциклопедия «Wikipedia» и поисковые системы Google, Yandex и Baidu.

Проект по идентификации аккаунтов пользователей социальной сети, придерживающихся здорового образа жизни и вегетарианства

1. Из всех сообществ «ВКонтакте» (~ 200 млн записей) находим полнотекстовым поиском те, которые в названии, описании или статусе имеют лингвистические маркеры. Маркерные слова разделены на две группы: тематика здорового образа жизни и тематика вегетарианства.

2. Полученные сообщества фильтруем по следующим критериям:

- 1) сообщество имеет 1000 и более подписчиков;
- 2) сообщество не является «событием»;
- 3) сообщество имеет посты с начала 2021 года.

Таким образом имеем три реестра сообществ:

- 1) здоровый образ жизни (2 996 сообществ);
- 2) вегетарианство (171 сообщество);
- 3) встретились маркеры обеих групп (247 сообществ).

3. Выгружаем подписчиков сообществ, проживающих в Москве или Санкт-Петербурге, с условием, что 25 и более процентов аудитории сообщества проживают в Москве либо Санкт-Петербурге (235 4050 записей).

4. На основе списка участников полученных сообществ выгружаем:

- 1) профили пользователей, которые были онлайн в течение месяца и не забанены (2 343 616 записей);
- 2) подписки пользователей (2 343 608 записей).

5. Выгружаем стены за год у сообществ, в которых от 25 и более процентов аудитории проживают в Москве либо Санкт-Петербурге (53 250 публикаций). Для всех публикаций выгружаем их лайки (43 999 записей) и комментарии (21 087 записей).

6. Подписчиков сообществ делим на 6 групп:

- a. Москвичи, подписанные на ЗОЖ сообщества.
- b. Москвичи, подписанные на вегетарианские сообщества.
- c. Москвичи, подписанные на оба типа сообществ.
- d. Петербуржцы, подписанные на ЗОЖ сообщества.
- e. Петербуржцы, подписанные на вегетарианские сообщества.
- f. Петербуржцы, подписанные на оба типа сообществ.

7. Для каждой из групп пользователей строим сводные таблицы:

- a. Профиль пользователя и список сообществ, на которые он подписан.
- b. Топ 30 подписок пользователей для каждого сочетания пола («не указан», «мужской», «женский») и возрастной группы («возраст не указан», «до 17», «18–35», «36–55», «56–80»).

8. Строим дополненную таблицу для выгруженных постов сообществ:

- a. Объект поста.
- b. Доли и количества подписчиков сообщества, проживающих в Москве и Санкт-Петербурге.
- c. Доли и количества комментариев и репостов, оставленных москвичами и петербуржцами.

В итоге имеем следующие таблицы:

- ✓ Профили сообществ: 3 414 уникальных записей, период выгрузки – сентябрь 2021.
- ✓ Посты сообществ за год: 53 250 уникальных записей, дата и время последнего поста – 2021-09-13 15:28:26.
- ✓ Лайки к постам: 43 999 уникальных записей, период выгрузки – сентябрь 2021.
- ✓ Комментарии к постам: 21 087 уникальных записей, период выгрузки – сентябрь 2021.
- ✓ Профили участников сообществ: 2 343 616 уникальных записей, период выгрузки – август 2021.
- ✓ Подписки участников сообществ: 2 343 608 уникальных записей, период выгрузки – сентябрь 2021.

Проект анализа радикализма в социальной сети

1. Из всех сообществ «ВКонтакте» (~ 200 млн записей) находим полнотекстовым поиском те, которые в названии, описании или статусе имеют лингвистические маркеры, связанные с тематикой радикализма и экстремизма.

2. Выявленные 169 сообществ группируем по следующим тематикам:

- 1) салафиты (39 сообществ);
- 2) праворадикалы (54 сообщества);
- 3) леворадикалы (76 сообществ).

3. Выгружаем контент, связанный с найденными сообществами:

- 1) списки участников сообществ;

- 2) профили незаблокированных участников, бывших онлайн в течение месяца на момент выгрузки (1 531 881 пользователь);
- 3) списки друзей участников (1 641 150 записей);
- 4) публикации сообществ за год (389 737 публикаций);
- 5) списки лайкнувших пользователей для каждой из публикаций (291 882 записи);
- 6) список комментариев с указанием ID автора для каждой из публикаций (54 369 записей);
- 7) публикации пользователей за год (109 696 590 публикаций).

4. На основе выгруженного контента создаём следующие таблицы для каждой тематической группы:

- 1) публикации пользователей, репостнутые из найденных сообществ;
- 2) количество лайков от пользователей из разных регионов по месяцам;
- 3) список самых активных участников, состоящих одновременно в нескольких сообществах тематической группы.

В итоге имеем следующие таблицы:

- ✓ Профили сообществ: 169 уникальных записей, период выгрузки – ноябрь 2020.
- ✓ Профили участников сообществ: 1 531 881 уникальная запись, период выгрузки – ноябрь 2020.
- ✓ Публикации сообществ за год: 389 737 уникальных записей, дата и время последнего поста – 2021-10-18 15:47.
- ✓ Лайки к публикациям сообществ: 291 882 уникальных записи, период выгрузки – ноябрь 2020, октябрь 2021.
- ✓ Комментарии к публикациям сообществ: 54 369 уникальных записей, период выгрузки – ноябрь 2020, октябрь 2021.
- ✓ Списки друзей пользователей: 1 641 150 уникальная запись, период выгрузки – октябрь 2020.
- ✓ Публикации подписчиков сообществ за год: 109 696 590 уникальных записей, период выгрузки – октябрь 2020.

Проект построения рейтинга университетов по критерию их популярности у пользователей интернет-энциклопедии Wikipedia и поисковых систем

Исследование учитывает три основных показателя: 1) статистика просмотров страниц университета в интернет-энциклопедии Wikipedia; 2) разработка и применение алгоритма поиска выдающихся выпускников университета с использованием данных Интернет-энциклопедии Wikipedia с применением методов обработки естественного языка; 3) разработка алгоритмов подсчета количества страниц веб-сайта университета, индексируемых поисковыми системами Google, Yandex и Baidu.

1. Имеется список университетов, предоставленный заказчиком. Таблица содержит ID университета, его название, страну и адрес официального сайта. По списку определяется перечень языков стран, университеты которых входят в список, для выгрузки страниц Wikipedia на соответствующих национальных языках. На конец 2021 г. имеется список из 1 977 университетов.

2. Выполняется полная выгрузка текстовой информации Wikipedia с помощью API (378 Гб данных). Выгрузка включает статьи на следующих языках: чешский, немецкий, английский, испанский, арабский, французский, ирландский, итальянский, японский, казахский, корейский, нидерландский, португальский, русский, шведский, турецкий, китайский.

3. Выполняется поиск страниц искомых университетов на английском и национальных языках. Поиск включает в себя автоматическую часть (поиск по текстовому шаблону) и ручную часть (ручная проверка данных и поиск недостающих университетов).

4. У найденных в Wikipedia страниц университетов выполняется подсчет количества просмотров на английском и национальном языках методами API.

5. Выполняется поиск выпускников университетов по всем выгруженным данным на национальном для страны университета и английском языках.

Алгоритм поиска включает в себя автоматизированные предобработку и постобработку данных лингвистическими методами с применением технологий обработки естественного языка (Natural Language Processing):

1) автоматическое извлечение ключевых слов и сущностей, статистическая обработка текста и разрешение кореференции. Иерархический поиск по ключевым словам частей статьи;

2) обработка предложений для удаления вводных конструкций для разбиения сложных предложений на простые и лемматизации;

3) определение субъекта обучения на основании дистанции слов в предложении;

4) удаление персон, имеющих определенное отношение к университету, но не являющихся его выпускниками (обучающиеся на текущий момент, отчисленные, сотрудники, выступающие за команду университета спортсмены и многое другое).

6. Формируется список выпускников университета, которым посвящена страница интернет-энциклопедии Wikipedia.

7. Определяется дата рождения каждого выпускника из списка на основе анализа конструкций метаразметки страниц интернет-энциклопедии Wikipedia.

8. Рассчитывается количество просмотров страниц идентифицированных выпускников университетов.

9. Используя API поисковых систем Google, Yandex и Baidu, выполняется подсчет количества страниц веб-сайтов университетов 4 раза с интервалом в 1 неделю для вычисления средних показателей.

10. Выполняется ручная проверка и «чистка» первых 5 000 выпускников, оказавшихся «наиболее влиятельными» (имеющих наибольшее количество просмотров).

5. Примеры наборов данных

В ходе прикладных и исследовательских проектов накапливаются данные. Однажды использованные данные (уже структурированные и обработанные) могут быть использованы повторно.

Тенденции «открытой науки» стимулируют политику открытого доступа к определённым данным, публикации датасетов, наличие репозиторий, обмен данными и др. Функционируют специальные типы журналов (Data Paper), публикующие наборы и описания данных.

Ниже приведено несколько примеров подготовленных в Центре наборов данных из социальной сети «ВКонтакте».

✓ Реестр российских официальных и неофициальных университетских сообществ «ВКонтакте»: 2 миллиона постов и комментариев, извлечённые со стен 2 168 сообществ 548 высших учебных заведений РФ и их филиалов (сентябрь 2020 – июль 2021).

✓ Официальные страницы «ВКонтакте» университетов, в составе которых функционируют 72 эндаумент-фонда высших учебных заведений; 1 369 тыс. профилей пользователей «ВКонтакте» – выпускников данных вузов (с идентификацией вуза в каждом случае).

✓ Профили 752 сообществ (ноябрь 2021), связанных с тематикой киберспорта; 5 645 159 профилей их участников (ноябрь 2021).

✓ Реестр 82 сообществ коренных малочисленных народов Севера, Сибири и Дальнего Востока; 32 616 постов; 53 085 профилей подписчиков данных сообществ (ноябрь 2021).

✓ Реестр сельских пользователей социальной сети «ВКонтакте» – 1 828 232 профиля, разделённые на 3 группы: «потенциально готовые к переезду в сельскую местность» – 251 452 профиля; «имеющие опыт проживания в сельской местности» – 1 676 602 профиля;

«проявляющие интерес к тематике сельской жизни» – группа оказалась на 100% принадлежащей второй части реестра – 53 652 профиля пользователей, имеющих опыт проживания в сельской местности и активно участвующих в обсуждениях публикаций по сельской тематике в соответствующих сообществах.

✓ Реестр благотворительных сообществ (2018/2021 – единый); посты (2018); сети благотворительных сообществ (2018); индексы активности благотворительных сообществ (2018).

✓ Реестр сообществ «ВКонтакте», посвящённых обсуждению вакцинации от COVID-19. Тематические сообщения о вакцинации от COVID-19 (~14 тыс. постов), сопровождаемые метаданными и размеченные по тематическим топикам.

✓ Реестр региональных сообществ с указанием их метаданных и цифровых следов пользователей (7 000 сообществ). Обучающая выборка для автоматической классификации сообщений по категориям качества жизни (100 тысяч сообщений, размеченных вручную). База сообщений отобранных региональных сообществ за 2019–2021 гг., автоматически классифицированных по категориям качества жизни с помощью алгоритма машинного обучения (8 млн сообщений). Индексы качества жизни населения по регионам РФ.

Заключение

Интернет-пользователи оставляют в сети цифровые следы, которые сохраняются и накапливаются. В настоящее время для аналитики в мире используется только небольшой процент данных.

Автоматизированный сбор данных открывает возможности для исследования социальных явлений на большом эмпирическом материале, в том числе в текущем временном срезе.

Потенциал больших данных должен использоваться в научных целях. Технологии их анализа и новые перспективы использования результатов в соответствии с задачами разных научных областей могут раскрыться только в исследовательской практике.

Список источников

1. Nayward G.M., Pearce L.D. The Sociology of Adolescent Religious and Spiritual Development // *Adolescent Research Review*. 2021. № 6 (3). P. 265–276.
2. Прима Я.Г., Прима П.А. Применение универсальных методов сбора данных в маркетинге и проектном менеджменте // *Экономические и социально-гуманитарные исследования*. 2018. № 3 (19). С. 57–70.
3. Мартышкин А.И., Сальников И.И., Пашенко Д.В. Этапы сбора и представления больших данных для построения социального профиля человека // *Известия Тульского государственного университета. Технические науки*. 2018. Вып. 9. С. 617–628.
4. Яцков А.К., Варламов М.И., Турдаков Д.Ю. Сбор и извлечение данных с веб-сайтов СМИ // *Программирование*. 2018. № 5. С. 68–80.
5. Нгуен Т.В., Кравец А.Г. Алгоритм работы веб-краулера для решения задачи сбора данных из открытых Интернет-источников // *Известия СПбГТИ (ТУ)*. 2019. № 51 (77). С. 115–119.
6. Поликарпов Е.С., Анисимов С.Л., Толстых А.А. О защищённости сайта сети Интернет от автоматизированного сбора данных // *Вестник Воронежского института МВД России*. 2020. № 1. С. 77–84.
7. Zavalin V.I., Miksa S.D. Collecting and Evaluating Large Volumes of Bibliographic Metadata Aggregated in the Worldcat Database: A Proposed Methodology to Overcome Challenges // *Electronic Library*. 2021. № 39 (3). P. 486–503.
8. Tisbeni S.R., Cesini D., Martelli B., Carbone A., Cavallaro C., Duma D.C., Falabella A., Galletti M., Gasparetto J., Furlan E., Michelotto D., Minarini F. A Big Data Platform for Heterogeneous Data Collection and Analysis in Large-scale Data Centres // *International Symposium on Grids and Cloud, ISGC 2021, Taipei, 22 March 2021 – 26 March 2021*. Vol. 37822 (October 2021), 0082021.
9. Tikito I., Souissi N. Smart Data Collection in Mobile Edge Computing Environment // *2020 International Conference on Intelligent Systems and Computer Vision, ISCV 2020, June 2020*, 9204277.
10. Реброва О.Ю., Федяева В.К. Мета-анализы и оценка их методологического качества. Русскоязычная версия вопросника AMSTAR // *Медицинские технологии. Оценка и выбор*. 2016. № 1 (23). С. 10–16.
11. Губа К. Большие данные в социологии: новые данные, новая социология // *Социологическое обозрение*. 2018. Т. 17, № 1. С. 213–236.

12. Shekhotin E., Myagkov M., Goiko V., Dunaeva D. Assessment of Quality of Life in Regions of Russia Based on Social Media Data // *Journal of Eurasian Studies*. 2021. Iss. 2, July. P. 1–17.
13. Палкин Р.В., Сапрыкин В.О., Гойко В.Л., Сайфулин Э.Р. VKAPI8. Библиотека методов по выгрузке и анализу данных из социальной сети «ВКонтакте». Свидетельство о регистрации программы для ЭВМ RU 2019662001, 13.09.2019. Заявка № 2019661005 от 10.09.2019.
14. Чудинов С.И., Сербина Г.Н., Мундриевская Ю.О. Сетевая организация скулшутеров в социальной сети «ВКонтакте» на примере фанатского сообщества «керченского стрелка» // *Мониторинг общественного мнения: экономические и социальные перемены*. 2021. № 4. С. 363–383.
15. Мацуга В.В., Гойко В.Л., Фещенко А.В., Уляков М.А., Петров Е.Ю. Платформа психологической диагностики с авторизацией респондентов в социальной сети «ВКонтакте». Свидетельство о регистрации программы для ЭВМ 2021612933, 26.02.2021. Заявка № 2021611585 от 15.02.2021.
16. Сербина Г.Н., Мацуга В.В., Гойко В.Л. Анализ связи психологических характеристик пользователей социальной сети «ВКонтакте» с подписками на сообщества с девиантным контентом // *Вестник Томского государственного университета*. 2021. № 467. С. 164–169.
17. Сайфулин Э.Р., Палкин Р.В., Сапрыкин В.О., Гойко В.Л., Чувашов И.Н., Благинин А.Л. База данных пользователей и сообществ социальной сети «ВКонтакте». Свидетельство о регистрации базы данных RU 2020621816, 08.10.2020. Заявка № 2020621672 от 30.09.2020.
18. Сайфулин Э.Р., Палкин Р.В., Сапрыкин В.О., Гойко В.Л., Коварж Г.Ю., Орлов С.А., Щекотин Е.В., Кашпур В.В., Пешковская А.Г. База данных региональных сообществ «ВКонтакте» Томской области. Свидетельство о регистрации базы данных RU 2019622274, 05.12.2019. Заявка № 2019622166 от 22.11.2019.
19. Мацуга В.В., Гойко В.Л., Фещенко А.В. База данных результатов психологической диагностики и пользовательских данных социальной сети «ВКонтакте» вузовской молодежи. Свидетельство о регистрации базы данных RU 2021621266, 15.06.2021. Заявка № 2021621135 от 07.06.2021.
20. Кашпур В.В., Барышев А.А., Чудинов С.И. Репрезентация радикальных сообществ в российских социальных медиа: специфика контента и индекс активности // *Вестник Томского государственного университета*. 2021. № 467. С. 133–143.
21. Митягина Е.В., Конышев Е.В., Чернышев К.А., Сайфулин Э.Р. Цифровые следы выпускников вузов при исследовании миграции из регионов-доноров // *Вестник Томского государственного университета*. 2021. № 467. С. 144–155.
22. Peshkovskaya A., Matsuta V., Evseev V. Time Spent Online on Social Networks Linked with Anxiety and Personality Traits // *European Neuropsychopharmacology*. 2020. Vol. 40. P. 387–388.
23. Kozitsina (Babkina) T., Goiko V., Khomutenko V., Palkin R., Mundrievskaya Yu., Myagkov M., Sukhareva M., Froumin I. Measuring University Impact: Wikipedia Approach // 2021 3rd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA), 10–12 November. 2021.

FROM THE EXPERIENCE OF ORGANIZING AUTOMATED DATA COLLECTION AT TOMSK UNIVERSITY

A.L. Blaginin, E.R. Sajfulin, A.Yu. Sarkisova

National Research Tomsk State University, Tomsk, Russian Federation

Abstract. The article systematizes methods, tools, sources, examples of the results of automated big data collection, updated in the practice of projects on socially significant topics in the Center for Applied Big Data Analysis of the National Research Tomsk State University in 2016–2022. The authors characterize (1) data extraction using the API, (2) web scraping, (3) emulation of the operation of a mobile application, (4) the use of specialized platforms, (5) the use of a database management system, (6) the obtaining of access to data directly from the owner of the Internet resource / data operator. The authors list the problems and limitations of automated data collection. Some examples of data sources are given. The social network VK is a source for which the Center currently has the most developed tools. The article contains examples of dataset descriptions and examples of the collection methodology. The article is intended for specialists in the field of social sciences and humanities. The objectives of the article are related to the popularization of relevant methods and resources for working with big data and the expansion of their use in research practice.

Keywords: automated data collection, big data, data collection, data sets, web scraping, open API, social networks, digital humanities

ПРОБЛЕМА ИЗМЕРЕНИЯ ВИРТУАЛЬНОЙ КУЛЬТУРНОЙ АКТИВНОСТИ В СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ» МЕТОДАМИ BIG DATA

Э.В. Бушкова-Шиклина, А.С. Москвин,
И.А. Кушова, М.П. Голенок, К.И. Одегова

Вятский государственный университет, Киров, Россия

Аннотация. Целью исследования явилась проблематизация измерения различных форм виртуальной активности (в частности, культурной активности) методами Big Data. Существует необходимость адаптации имеющихся методик/индексов измерения виртуальной активности. Авторами представлен расчет Индекса культурной активности пользователей (ИКАП) и «Индекса культурной активности администраторов сообществ» (ИКАС) в социальной сети «ВКонтакте» с использованием методов Big Data. На основании эмпирического исследования массива данных социальной сети представлены различия виртуальной культурной активности в разрезе регионов (Кировская область, Нижегородская область, Республика Татарстан), направлений активности (культурно-просветительское направление, театрално-зрелищное направление, дополнительное внешкольное образование, рекреационно-досуговое направление, религиозные сообщества, а также сообщества, посвященные здоровому образу жизни), статуса паблика (официальный и неофициальный). Показано, что наиболее активными во взаимодействии с культурными пабликами являются подписчики Республики Татарстан. В целом по выборке наибольшая активность подписчиков отмечается по рекреационно-досуговому направлению и направлению здорового образа жизни. Подавляющее большинство всех анализируемых культурных пабликов являются официальными страницами различных организаций. Официальные сообщества ведут более организованную работу с контентом и подписчиками, численность их пользователей в два раза выше численности неофициальных пабликов.

Ключевые слова: индекс культурной активности пользователей, индекс культурной активности администраторов сообществ, социальная сеть, большие данные

Введение

Исследование социально-экономических и культурных сторон общественной жизни при помощи больших данных стало устойчивой тенденцией в последние годы. Данный тренд можно проследить в исследованиях О.В. Крыштановской, К. Губы, А.В. Одинцова, И.Б. Орловой и Е.В. Фомина, Е.В. Щекотина [1–6] и др. Значительное внимание в этом ключе уделяется анализу доступных источников больших данных – социальных сетей. В связи с этим многие работы посвящены изучению социальных и социально-психологических явлений посредством анализа социальных сетей (В.А. Ермаков, Г.Т. Журавлев, Е.В. Ковалевская, В.Г. Писаревский, А.Р. Гапич, С.С. Асатрян, О.В. Минкина, Л.А. Забокрицкая, Н.А. Хлебников, Т.А. Орешкина, Е.И. Комоцкий [7–10] и др.).

Все чаще предметом анализа становится категория социальной активности пользователей социальных сетей – виртуальной активности (например, в работах Л.А. Андреева, А.С. Садчиковой, А.Г. Поляковой, Э.М. Ахметшина [11–12] и др.). Частыми являются работы, связанные с изучением виртуальной активности молодежи, виртуальной игровой активности [13]; сравниваются виртуальная и реальная социальная активность [14], а также исследуются различные формы девиантной виртуальной активности [15].

Категория *виртуальной активности* рассматривается в литературе как производная форма социальной активности, которая, в свою очередь, понимается как мера социальной преобразующей деятельности в рамках одной или нескольких сфер общественной жизни (трудовой, политической, познавательно-творческой и пр.), основанная на осознании внутренней необходимости действий в соответствии с индивидуальными и общественными по-

требностями [16]. С позиции психологии виртуальная активность рассматривается как диалектическое взаимодействие операционально-технической и смысловой сфер сознания (деятельности) человека, за счет чего происходит развитие личности и ее сознания в новой виртуальной реальности [17].

Отдельной становится проблема *измерения* виртуальной активности. Конечно, методы и методики измерения виртуальной активности зависят от сферы ее проявления. Например, разрабатываются специальные методики изучения девиантной виртуальной активности [18], лингвистического анализа сообщений для безопасности виртуального общения [19] и пр. Актуальной становится проблема измерения методами Big Data особенностей интернет-аудитории и пользовательской активности [20]. Разрабатываются так называемые индексы: виртуальности, вовлеченности. Например, предлагается *индекс виртуальности*, включающий показатели immersion (погруженности), involvement (вовлеченности), interactivity (интерактивности) [21]. Погруженность характеризует охваченность органов чувств человека, находящегося в искусственно созданной среде. Вовлеченность указывает на рациональную и эмоциональную составляющие внутреннего мира человека. Интерактивность определяет процессы взаимодействия пользователя и виртуальной среды.

Более частным является *индекс вовлеченности*, рассматриваемый в зависимости от практической задачи как оценка реакции на виртуальный продукт, процесс взаимодействия пользователя с сайтом либо наиболее широко – как наличие эмоциональной, когнитивной и поведенческой связи между пользователем и ресурсом в процессе их взаимодействия [22–23]. Отмечается четыре подхода к измерению виртуальной вовлеченности: самоотчет, физиологический подход, нейрофизиологический подход и веб-аналитика. Именно последний подход позволяет задействовать большие данные сайтов и социальных сетей, так как оперирует такими доступными для анализа онлайн-поведенческими показателями, как рейтинг кликов, количество просмотренных страниц, время онлайн-просмотра, частота повторных посещений и пр. [23].

Таким образом, анализ теоретических источников показал, что проблема измерения различных форм виртуальной активности является актуальной и перспективной для дальнейших разработок. Виртуальная активность – частный случай социальной активности различных групп населения. Методы ее оценки весьма разнообразны, зависят от формы активности и используемых интернет-ресурсов. Перспективным становится разработка различных индексов виртуальной активности – индексов виртуальности, вовлеченности и пр. Одним из наиболее доступных источников анализа виртуальной культурной активности являются открытые социальные сети, позволяющие задействовать методы Big Data и веб-аналитику.

Методы исследования

В данной статье представлен анализ виртуальной культурной активности пользователей социальной сети ВКонтакте в следующих аспектах:

1) *региональном* – в разрезе пабликов, имеющих региональную привязку к Кировской и Нижегородской областям, а также Республике Татарстан;

2) *по направлениям культурной активности*: культурно-просветительское (музеи, клубы, библиотеки, выставочные залы); театрално-зрелищное (театры, концертные организации, цирки, кинотеатры); дополнительное внешкольное образование (музыкальное, хореографическое, художественные школы, подготовка к школе); рекреационно-досуговое (центры и базы отдыха, хобби, парки развлечений и т.д.); религиозные сообщества; здоровый образ жизни (здоровое питание, спортивные секции, фитнес);

3) в зависимости от *статуса сообщества*: официальное представительство культурной организации или неофициальное сообщество.

В рамках данного исследования проводился анализ сообществ, *критериями отбора* которых были: 1) последняя запись не позднее марта 2022 г.; 2) подписчиков не менее 200 чел.; 3) имеется региональная привязка в названии и/или описании и/или в адресе/по фильтру в

«ВКонтакте». В результате был получен реестр сообществ по культуре Кировской и Нижегородской областей, Республики Татарстан (всего 1 289 сообществ) и реестр постов за период с 01.10.2021 по 03.04.2022 (всего 106 322 поста).

Результаты и обсуждение

Для анализа виртуальной активности разработаны Индексы культурной активности пользователей (ИКАП) и Индекс культурной активности организаторов сообществ (ИКАС). Представим их формулы.

Методика расчета Индекса культурной активности пользователей (ИКАП).

В данной методике рассчитывается среднее взвешенное значение четырех показателей активности пользователей на публикуемые посты (табл. 1), далее рассчитывается индекс ИКАП (1):

$$\begin{aligned} \text{ИКАП} &= (\omega_1 \dot{x}_{\text{пр}} + \omega_2 \dot{x}_{\text{л}} + \omega_3 \dot{x}_{\text{р}} + \omega_4 \dot{x}_{\text{к}}) / 4; \\ \text{ИКАП} &= (0,1 \dot{x}_{\text{пр}} + 0,2 \dot{x}_{\text{л}} + 0,3 \dot{x}_{\text{р}} + 0,4 \dot{x}_{\text{к}}) / 4. \end{aligned} \quad (1)$$

Т а б л и ц а 1

Составляющие ИКАП

Уровень активности	Вид активности	Компонента индекса: средняя	Вес средней, ω	Индекс ИКАП
Низкий	Просмотр	$\dot{x}_{\text{пр}}$	0,1	$(0,1 \dot{x}_{\text{пр}} + 0,2 \dot{x}_{\text{л}} + 0,3 \dot{x}_{\text{р}} + 0,4 \dot{x}_{\text{к}}) / 4$
Средний	Лайк	$\dot{x}_{\text{л}}$	0,2	
Средний	Репост	$\dot{x}_{\text{р}}$	0,3	
Высокий	Комментарий	$\dot{x}_{\text{к}}$	0,4	

Методика расчета Индекса культурной активности сообществ (ИКАС). В рамках данной методике рассчитывается среднее взвешенное значение семи показателей активности администраторов сообществ (табл. 2), далее рассчитывается индекс ИКАС (2):

$$\begin{aligned} \text{ИКАС} &= (\omega_1 \dot{x}_{\text{пп}} + \omega_2 \dot{x}_{\text{пст}} + \omega_3 \dot{x}_{\text{в}} + \omega_4 \dot{x}_{\text{ф}} + \omega_5 \dot{x}_{\text{а}} + \omega_6 \dot{x}_{\text{ст}} + \omega_7 \dot{x}_{\text{м}}) / 7; \\ \text{ИКАС} &= (0,21 \dot{x}_{\text{пп}} + 0,24 \dot{x}_{\text{пст}} + 0,01 \dot{x}_{\text{в}} + 0,01 \dot{x}_{\text{ф}} + 0,01 \dot{x}_{\text{а}} + 0,01 \dot{x}_{\text{ст}} + 0,01 \dot{x}_{\text{м}}) / 7. \end{aligned} \quad (2)$$

Т а б л и ц а 2

Составляющие ИКАС

Вид активности	Компонента индекса: средняя	Вес средней, ω	Индекс ИКАП
Посещаемость сообщества, кол-во подписчиков	$\dot{x}_{\text{пп}}$	0,21	$(0,21 \dot{x}_{\text{пп}} + 0,24 \dot{x}_{\text{пст}} + 0,01 \dot{x}_{\text{в}} + 0,01 \dot{x}_{\text{ф}} + 0,01 \dot{x}_{\text{а}} + 0,01 \dot{x}_{\text{ст}} + 0,01 \dot{x}_{\text{м}}) / 7$
Посты за определенный период, ед.	$\dot{x}_{\text{пст}}$	0,24	
Видеофайлы, ед.	$\dot{x}_{\text{в}}$	0,01	
Фотоальбомы, ед.	$\dot{x}_{\text{ф}}$	0,01	
Аудиофайлы, ед.	$\dot{x}_{\text{а}}$	0,01	
Статьи, ед.	$\dot{x}_{\text{ст}}$	0,01	
Мероприятия, ед.	$\dot{x}_{\text{м}}$	0,01	

Исходя из анализа ИКАП (табл. 3), можно сделать следующие выводы.

1. В целом по выборке наиболее высокий ИКАП наблюдается в сообществах Республики Татарстан (ИКАП = 68), что в 2–3 раза выше, нежели в Нижегородской и Кировской областях и в целом по всей выборке (ИКАП = 39). В целом по всему массиву доминирует активность подписчиков по направлению здорового образа жизни, а также по рекреационно-досуговому направлению (ИКАП = 107 и 50 соответственно), наименьшая активность отмечается по культурно-просветительскому и дополнительному образовательному направлениям (ИКАП = 15 и 18 соответственно).

2. В Кировской области наибольшая виртуальная активность наблюдается по рекреационно-досуговому направлению (ИКАП = 42), а также направлениям здорового образа жизни

и театральнo-зрелищнoму (35 и 32). В Нижегородской области аналогичная ситуация (ИКАП = 64, 37 и 47 соответственно). В Республике Татарстан ситуация аналогичная, но активность по направлению здорового образа жизни значительно доминирует (ИКАП = 225), на втором месте – показатели высокой активности в религиозных сообществах (ИКАП = 66).

Т а б л и ц а 3

**ИКАП в социальной сети ВКонтакте
(за период 10.2021–03.2022 г.)**

Направление	Просмотры, среднее	Лайки, среднее	Репосты, среднее	Коммен- тарии, среднее	ИКАП
По выборке	1501	28	3	2	39
Культурно-просветительское	544	16	1	1	15
Театральнo-зрелищнoе	1420	18	3	3	37
Дополнительное внешкольное образование	708	11	2	1	18
Рекреационно-досуговое	1960	19	4	2	50
Религиозные сообщества	1272	48	4	1	35
Здоровый образ жизни	4092	73	4	8	107
Кировская обл.	1156	18	3	1	30
Культурно-просветительское	702	18	2	1	19
Театральнo-зрелищнoе	1250	14	3	2	32
Дополнительное внешкольное образование	947	13	2	1	24
Рекреационно-досуговое	1642	18	4	2	42
Религиозные сообщества	1156	37	3	0	31
Здоровый образ жизни	1323	29	3	1	35
Нижегородская обл.	904	24	2	1	24
Культурно-просветительское	304	17	1	0	8
Театральнo-зрелищнoе	1786	40	4	5	47
Дополнительное внешкольное образование	527	10	2	0	14
Рекреационно-досуговое	2452	36	6	2	64
Религиозные сообщества	919	37	3	1	25
Здоровый образ жизни	1429	17	1	4	37
Республика Татарстан	2602	46	4	4	68
Культурно-просветительское	502	12	1	1	13
Театральнo-зрелищнoе	1536	17	3	3	40
Дополнительное внешкольное образование	272	3	0	1	7
Рекреационно-досуговое	2483	16	4	3	63
Религиозные сообщества	2414	85	9	3	66
Здоровый образ жизни	8591	162	7	14	225

Исходя из анализа ИКАС (табл. 4), можно сделать следующие выводы.

1. В целом по всем регионам отмечается примерно одинаковый уровень виртуальной активности сообществ (ИКАС = 1262–1498). При этом доминирует активность культурно-просветительских и рекреационно-досуговых сообществ (ИКАС = 1010 и 913 соответственно).

2. Наибольшая активность сообществ Кировской области и Республики Татарстан отмечается по рекреационно-досуговым и театральнo-зрелищным направлениям (ИКАС = 581, 544; 667, 475 соответственно), в Нижегородской области – по религиозному (ИКАС = 486).

Представим результаты статистического анализа виртуальной культурной активности пользователей и сообществ социальной сети «ВКонтакте». Выборка исследования составила 1 289 сообществ, в каждом регионе – в относительно одинаковых пропорциях 28–36% (рис. 1).

**Индекс культурной активности сообществ (ИКАС) в социальной сети ВКонтакте
(за период 10.2021 – 03.2022)**

Направление	Количество участников, среднее	Количество постов, в мес.	Видео, среднее	Фотоальбомы, среднее	Аудио, среднее	Статьи, среднее	Мероприятия, среднее	ИКАС
По выборке	4132,9	177200	109	69,3	24,3	10,6	5,3	6200
Культурно-просветительское	2162,4	27570	83,8	45,2	9,3	27,3	7,6	1010
Театрально-зрелищное	9493,7	14952	259,4	83	66,1	12,9	10,4	798
Дополнительное внешкольное образование	2215,6	14497	80	29,7	11,6	8	0,8	564
Рекреационно-досуговое	8384,9	19293	132,8	107,8	50,1	6,8	18,4	913
Религиозные сообщества	4418,7	16838	155,4	101,9	100,3	32,7	0,4	710
Здоровый образ жизни	2716,2	13171	75,9	66,2	5,7	5,4	1,7	533
Кировская обл.	3432,4	40689	105	54,2	22,1	9,1	4,3	1498
Культурно-просветительское	2503,5	11883	80,4	45,6	12	30,9	13,1	483
Театрально-зрелищное	9822,5	7251	320,3	67,7	19,6	20,4	5,7	544
Дополнительное внешкольное образование	1863	6644	60,4	25,8	4,3	6,4	0,4	284
Рекреационно-досуговое	5682,8	11948	151,2	69,2	100,5	5,7	11,7	581
Религиозные сообщества	1858,7	1092	80,9	180,5	7	12,5	0,1	94
Здоровый образ жизни	2878,4	1871	90,7	49,7	4,1	4,1	0,7	151
Нижегородская обл.	3218,8	34270	96,6	70,4	28,2	10,8	8,2	1272
Культурно-просветительское	1643,4	6242	126,9	24,9	18	66,4	5,5	264
Театрально-зрелищное	10484,9	1342	261,9	53,4	161	6,3	17,6	361
Дополнительное внешкольное образование	2097	7230	57,9	30,3	9	4,2	1,1	311
Рекреационно-досуговое	5519,2	1141	155,8	109,5	24,7	10,8	35,9	205
Религиозные сообщества	2532,5	11941	79,5	114,7	83,7	26,4	0,5	486
Здоровый образ жизни	1995,2	6374	66,1	65,5	6,6	6,8	2,3	279
Республика Татарстан	6216,6	31362	130,3	87,4	22,2	12,4	1,4	1262
Культурно-просветительское	1988,5	9445	76,6	49,5	5	14,9	0,6	384
Театрально-зрелищное	8554,7	6359	230,3	113,5	10,8	14,9	0,1	475
Дополнительное внешкольное образование	2575,3	623	108,5	32,3	19,1	11,4	1,2	99
Рекреационно-досуговое	15145,6	6204	82,7	156,1	13,4	3,8	0,6	667
Религиозные сообщества	9391,5	3805	337,1	25,6	189	56,8	0	413
Здоровый образ жизни	4571,6	4926	65,4	114,9	7,1	4,6	2,9	306

В среднем по выборке количество подписчиков в сообществе составило 4 133 человек, медианное значение – 1 421 человек (медиана в пабликах Республики Татарстан выше, чем в нижегородских и кировских – в среднем на 600–700 человек), минимальное количество подписчиков – 17 человек, максимальное – 200 115 человек (табл. 5, рис. 2).

Подавляющее большинство пабликов являются официальными страницами различных организаций (1 133 ед., 89%), среднее количество подписчиков в официальных сообществах практически в два раза выше, нежели в неофициальных сообществах (4 323 и 2 754 человек соответственно).

Около половины (45%) пабликов посвящены здоровому образу жизни, менее всего (5,8%) принадлежит религиозным сообществам (табл. 6).



Рис. 1. Соотношение количества сообществ в выборке исследования

Таблица 5

Статистика наполняемости сообществ социальной сети «ВКонтакте»

Регион	Среднее	Медиана	Минимум	Максимум
Кировская область	3 432	1 293	205	200 115
Нижегородская область	3 219	1 376	200	79 560
Республика Татарстан	6 217	1 941	17	152 663
По выборке	4 133	1 421	17	200 115



Рис. 2. Среднее количество пользователей в социальной сети «ВКонтакте»

Таблица 6

Структура тематики пабликов

Направление	%
Культурно-просветительское	11,1
Театрально-зрелищное	7,8
Дополнительное внешкольное образование	13,5
Рекреационно-досуговое	16,0
Религиозное	5,8
Здоровый образ жизни	44,7
Комплексное	1,1
Всего	100

Во всех регионах доминирует доля сообществ по направлению здорового образа жизни (22,8–56,6%), а также рекреационно-досуговому направлению (14–17%). В сообществах Республики Татарстан распределение сообществ по тематикам более сбалансированное: доля каждого направления колеблется в пределах 12–22% (χ^2 -квadrat Пирсона $\chi^2 = 165$, $df = 12$, $P = 0,0$) (рис. 3).

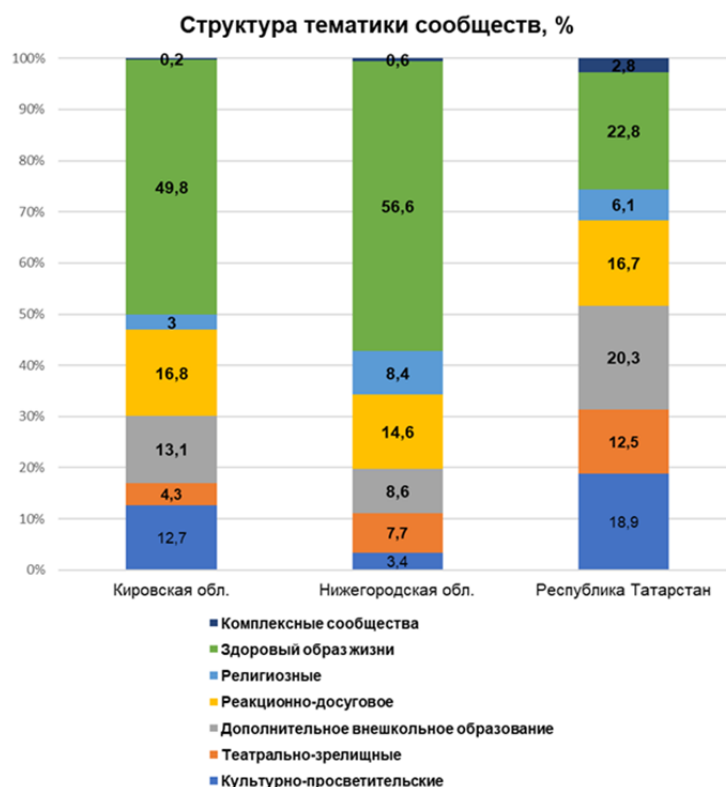


Рис. 3. Структура тематики сообществ в разрезе регионов

В официальных и неофициальных сообществах преобладает контент здорового образа жизни; структура неофициальных пабликов более сбалансирована: пользователи отдают предпочтение театрально-зрелищному и дополнительному дошкольному образованию (см. рис. 4).

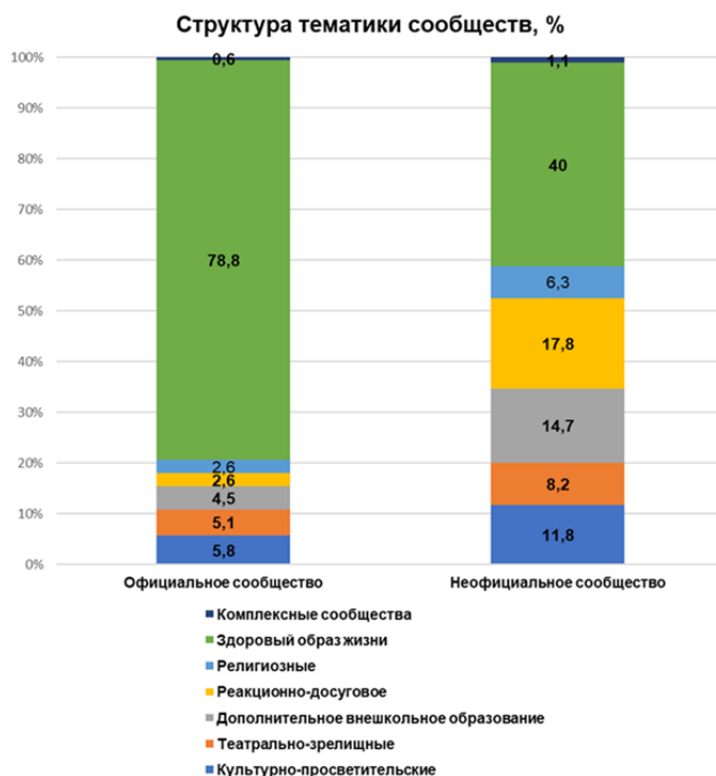


Рис. 4. Структура тематики официальных и неофициальных сообществ

Выводы

Целью данной статьи явилась проблематизация измерения различных форм виртуальной активности (в частности, культурной активности) методами Big Data.

Показано, что виртуальная культурная активность – частный случай социальной активности, для удовлетворения разнообразных культурных потребностей населения – рекреационно-досуговых, познавательно-творческих, потребностей в дополнительном внешкольном образовании и пр.

Анализ виртуальной культурной активности методами Big Data имеет методологические и методические ограничения, в частности, имеется необходимость адаптации имеющихся методик/индексов измерения виртуальной активности. В рамках данной статьи представлен расчет ИКАП и ИКАП.

При помощи указанных индексов, а также путем статистического анализа показаны особенности культурной активности пользователей и организаторов сообществ в разрезе регионов (Кировская область, Нижегородская область, Республика Татарстан), направлений активности (культурно-просветительское направление, театральное-зрелищное направление, дополнительное внешкольное образование, рекреационно-досуговое направление, религиозные сообщества, а также сообщества, посвященные здоровому образу жизни), статуса паблика (официальный и неофициальный).

Показано, что наиболее активными во взаимодействии с культурными пабликами являются подписчики Республики Татарстан, в основном, по направлению Здоровый образ жизни и религиозному направлению. В целом по всему массиву наибольшая активность подписчиков отмечается по рекреационно-досуговому направлению и направлению здорового образа жизни. При этом виртуальная активность организаторов культурных сообществ примерно одинаковая во всех регионах. При этом доминирует активность культурно-просветительских и рекреационно-досуговых сообществ. Наибольшая активность сообществ Кировской области и Республики Татарстан отмечается по рекреационно-досуговым и театральное-зрелищным направлениям, в Нижегородской области – по религиозному.

Подавляющее большинство (89%) анализируемых культурных пабликов являются официальными страницами различных организаций. Среднее количество подписчиков в официальных сообществах практически в два раза выше, нежели в неофициальных сообществах и составляет более четырех тысяч человек, структура контента по указанным направлениям в официальных сообществах более сбалансированная, нежели в неофициальных.

Список источников

1. Крыштановская О.В. Бесконтактная социология: новые формы исследования в цифровую эпоху // Цифровая социология. 2018. № 1. С. 4–8.
2. Губа К. Большие данные в социологии: новые данные, новая социология? // Социологическое обозрение. 2018. № 1, т. 17. С. 213–236.
3. Одинцов А.В. Открытость баз данных как условие формирования «больших данных» в социологии // Научно-методический электронный журнал «Концепт». 2017. № 12. С. 101–106.
4. Одинцов А.В. Социология общественного мнения и вызовы Big Data // Мониторинг общественного мнения: Экономические и социальные перемены. 2017. № 3. С. 30–43.
5. Орлова И.Б., Фомин Е.В. Цифровая социология: возможности, риски, перспективы // Национальная безопасность. 2020. № 3. С. 48–63.
6. Щекотин Е.В. Цифровые следы как новый источник данных о качестве жизни и благополучии: обзор современных тенденций // Вестник Томского государственного университета. 2021. № 467. С. 170–181.
7. Ермаков В.А., Журавлев Г.Т., Ковалевская Е.В. Анализ активности студентов в социальных сетях Интернета // Интерактивная наука. 2016. № 8. С. 44–48.
8. Писаревский В.Г. Аудитория социальных сетей и Интернет-сообщества как новые направления «Цифровой социологии» // Фундаментальные и прикладные исследования: проблемы и результаты. 2013. № 6. С. 96–98.
9. Гапич А.Э., Асатрян С.С., Минкина О.В. Визуальный контент социальных сетей и молодежный радикализм: основные методологические подходы и концептуальный дизайн исследования // Гуманитарные, социально-экономические и общественные науки. 2021. № 12/1. С. 38–42.

10. Забокрицкая Л.Д., Хлебников Н.А., Орешкина Т.А., Комоцкий Е.И. Возможности изучения ценностей молодежи через профиль социальной сети «ВКонтакте» // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 2. С.148–167.
11. Андреев А.Л., Садчикова А.С. Поисковые запросы в Интернете как средство анализа социального поведения россиян в условиях социально-экономического кризиса (на примере Приволжского федерального округа) // Социологическая наука и социальная практика. 2016. № 4 (16). С. 7–18.
12. Полякова А.Г., Ахметшин Э.М. Цифровой мониторинг социально-экономического положения населения города // Урбанистика. 2020. № 1. С. 148–154.
13. Дутов А. Виртуальная игровая активность московских учащихся (на примере студентов вузов и учащихся старших классов) // COLLEGIUM LINGUISTICUM – 2019 : тезисы докладов Ежегодной конференции Студенческого научного общества МГЛУ. М., 2019. С. 216.
14. Сунгурова Н.Л., Акимкина Ю.Е. Виртуальная и реальная социальная активность студентов // Ананьевские чтения – 2021 : материалы международной научной конференции / под общ. ред. А.В. Шаболтас ; отв. ред. В.И. Прусаков. СПб., 2021. С. 211–212.
15. Воробьева К.А. Виртуальная девиантная активность подростков: причины, формы, профилактика // Мир образования – образование в мире. 2021. № 4 (84). С. 184–194.
16. Завгородний М.Д. Социальная активность в сети Интернет // Вестник Российского университета дружбы народов. Серия: Социология. 2012. № 2. С. 49–55.
17. Овчинникова Т.Н. Виртуальная активность и ее особенности // Инновационные технологии в кинематографе и образовании : материалы и доклады VIII Международной научно-практической конференции. М., 2022. С. 174–186.
18. Шаров А.А. Методика изучения девиантной активности в реальной и виртуальной среде // Известия Саратовского университета. Новая серия. Серия: Акмеология образования. Психология развития. 2019. Т. 8, № 1. С. 30–37.
19. Никольская В.А., Родькина О.Я. Методы лингвистического анализа публичных текстовых сообщений в социальных сетях и возможности их применения для обеспечения безопасности // Успехи современной науки и образования. 2017. № 9. С. 160–163.
20. Варганов С.А. Большие данные в онлайн-СМИ: подходы и стратегии использования // Медиаскоп. 2017. № 4. С. 3.
21. Елхова О.И. Индекс виртуальности: философское обоснование // Вестник Северного (Арктического) федерального университета. Серия: Гуманитарные и социальные науки. 2021. Т. 21, № 3. С. 99–107.
22. Фролов А.А., Агурова А.А. Индексный анализ гражданской активности в социальных сетях // Известия Иркутского государственного университета. Серия Политология. Религиоведение. 2019. Т. 29. С. 29–43.
23. Воронина А.А., Шабалина О.А., Катаев А.В. Методы оценки вовлеченности пользователей интерактивных приложений // ИТНОУ: информационные технологии в науке, образовании и управлении. 2019. № 4. С. 70–74.

THE PROBLEM OF MEASURING VIRTUAL CULTURAL ACTIVITY IN THE SOCIAL NETWORK VKONTAKTE USING BIG DATA METHODS

E.V. Bushkova-Shiklina, A.S. Moskvina, I.A. Kushova, M.P. Golenok, K.I. Odegova

Vyatka State University, Kirov, Russian Federation

Abstract. The aim of the article was to problematize the measurement of various forms of virtual activity (in particular, cultural activity) using Big Data methods. There is a need to adapt existing methods/indices for measuring virtual activity. The authors present the calculation of the Index of Users' Cultural Activity and the Index of Community Administrators' Cultural Activity in the social network VKontakte using Big Data methods. Based on an empirical study of the social network data array, differences in virtual cultural activity are presented in the context of regions (Kirov Region, Nizhny Novgorod Region, Republic of Tatarstan), areas of activity (cultural and educational direction, theatrical and entertainment direction, additional extracurricular education, recreational and leisure direction, religious communities, as well as communities dedicated to a healthy lifestyle), public status (official and unofficial). The authors show that the subscribers of the Republic of Tatarstan are the most active in interaction with cultural public sites. In general, according to the sample, the greatest activity of subscribers is noted in the recreational direction and in the direction of a healthy lifestyle. The vast majority of all analyzed cultural publications are the official pages of various organizations. Official communities conduct more organized work with content and subscribers; the number of their users is twice as high as the number of unofficial public pages.

Keywords: index of users' cultural activity, index of community administrators' cultural activity, social network, big data

ИССЛЕДОВАНИЕ УРОВНЯ ТРЕВОЖНОСТИ, СВЯЗАННОГО С COVID-19: ВЛИЯНИЕ СОЦИАЛЬНО-ДЕМОГРАФИЧЕСКИХ И ЛИЧНОСТНЫХ ХАРАКТЕРИСТИК

С.А. Галкин^{1, 2}, Д.Г. Найденко², В.Л. Гойко²

¹ Томский национальный исследовательский медицинский центр Российской академии наук, Томск, Россия

² Национальный исследовательский Томский государственный университет, Томск, Россия

Аннотация. Проведена оценка уровня тревожности, связанного с COVID-19, а также влияние на него социально-демографических и личностных характеристик. В ходе онлайн-тестирования было опрошено 230 лиц в возрасте 18–49 лет. Обнаружено, что у 50 (21,74%) респондентов отмечается наличие дисфункциональной тревоги, связанной с коронавирусом по опроснику CAS. Также было обнаружено, что на уровень тревоги, связанной с COVID-19, оказывали влияние смертность среди близких от коронавируса ($r = 0,44$; $p = 0,045$), а также некоторые личностные особенности респондентов: уровень реактивной тревожности ($r = 0,36$; $p < 0,001$), поиск социальной поддержки ($r = 0,25$; $p = 0,03$), избегание проблем ($r = 0,27$; $p < 0,001$), экстраверсия ($r = -0,30$; $p < 0,001$), доброжелательность ($r = -0,27$; $p < 0,001$), добросовестность ($r = -0,29$; $p < 0,001$) и нейротизм ($r = 0,43$; $p < 0,001$). Полученные данные могут быть полезны при разработке программ, с целью уменьшить тревожность и страх общества перед коронавирусом.

Ключевые слова: новая коронавирусная инфекция, COVID-19, тревога, страх, CAS, личностные особенности

Источник финансирования: исследование выполнено при поддержке РФФИ, проект № 20-04-60301.

Введение

Новая коронавирусная болезнь (COVID-19) была объявлена общемировой пандемией Всемирной организацией здравоохранения (ВОЗ) в марте 2020 г., но первая ее вспышка была зарегистрирована в Китае еще в декабре 2019 г. Число подтвержденных случаев на февраль 2022 г. составляет более 414 млн человек во всем мире, в то время как число погибших от COVID-19 превысило 5,8 млн (статистика по миру на основе данных ВОЗ по состоянию на 16 февраля 2022 г.) [1].

Пандемия COVID-19, несомненно, привела к серьезнейшим экономическим, политическим, социальным и психологическим последствиям. В то время как первоначально внимание исследователей и врачей было направлено на физические проявления инфекции (например, пневмонию, лихорадку, кашель, повышенную утомляемость, одышку, головную боль и боль в горле) [2], научное сообщество вскоре выявило негативное влияние инфекции и на другие системы. Сейчас активно обсуждается влияние COVID-19 на головной мозг [3, 4] и психическое здоровье [2, 5–7]. Особую роль здесь играет возраст. Установлено, что вероятность летального исхода у пожилых больных с COVID-19 значительно выше, чем у молодых [2]. Однако, с точки зрения влияния COVID-19 на психологическое восприятие и психическое здоровье все не так однозначно. Исследование, проведенное A. Germani et al. (2020) [8], показало, что молодые люди больше беспокоятся и боятся пандемии COVID-19, чем старшее поколение, из-за большей вероятности стать потенциальными симптоматическими или бессимптомными носителями, что может способствовать большему распространению заболеваемости. В рекомендациях Межведомственного комитета по охране психического здоровья и психосоциальной поддержке рекомендуется интегрировать несколько уровней вмешательства в ответ на пандемию [9]. В этом контексте исследователи и клинические психологи пытаются устранить такие последствия, как тревожность, стресс и беспокойство, тщательно проверяя психосоциальное воздействие COVID-19, и соответствующим образом планируя конкретные индивидуальные терапевтические вмешательства. В настоящее время исследова-

тели также пытаются разработать методы, направленные на предотвращение развития психологического дистресса, связанного с COVID-19 [10].

Фактически, за последние несколько месяцев среди населения во всем мире наблюдалось значительное увеличение психологического дистресса и связанных с ним расстройств (например, генерализованной тревоги, панического расстройства и острого стрессового расстройства) [11, 12]. Например, T. Iachini, 2020 [11] исследовал психологические факторы, влияющие на соблюдение социальной дистанции (1,5 м согласно рекомендациям ВОЗ) и его влияние на психологическое благополучие, и обнаружил, что более высокие уровни тревоги, страха и стресса прямо коррелировали с соблюдением социальной дистанции, но при этом вызывая значительное снижение психологического самочувствия. Тревога и страх в условиях постоянного стресса, особенно во время пандемии COVID-19, приводят не только к формированию неуверенности, отрицания, бессонницы, диссоциативных симптомов, депрессивных расстройств, эмоционального дистресса, суицидальных мыслей и попыток, злоупотреблению алкоголем и психоактивными веществами, но и значительно повышают риск рецидива у людей с ранее существовавшими психическими и поведенческими расстройствами [6, 13–15].

Отрицательные эмоции, такие как тревога и страх, могут быть связаны с рядом аспектов пандемии COVID-19, в частности, с физическим ухудшением самочувствия, длительным стрессом, смертью близкого человека, социальной изоляцией, финансовыми трудностями и неуверенностью в будущем [16]. Действительно, исследование, проведенное Y. Huang и N. Zhao (2020) [16], показало, что общий уровень тревожности среди населения в целом составлял примерно 35,1% в период первой вспышки пандемии COVID-19. Эти данные имеют особое значение и свидетельствуют о важности приоритизации психологического тестирования, направленного на углубленное понимание признаков и симптомов стресса, связанного с COVID-19, а также на разработку адекватных методов диагностики и своевременных вмешательств.

Целью этого исследования было изучение уровня тревожности, связанного с COVID-19, с помощью Шкалы тревоги по поводу коронавируса (Coronavirus Anxiety Scale – CAS) [17, 18], среди жителей России. В частности, было проведено исследование на предмет того, влияют ли стабильные личностные характеристики (особенности личности) и социально-демографические факторы на тревогу и страх, связанные с COVID-19, и могут ли определенные условия, такие как положительный результат на COVID-19 или заражение среди родственников и друзей, еще больше усилить эти симптомы?

Материал и методы

Участники. В связи с необходимостью соблюдать самоизоляцию, исследование проводилось в формате онлайн-тестирования на платформе *ivik.org*. В ходе онлайн-тестирования, проведенного с января по апрель 2021 г., было опрошено 230 лиц в возрасте 18–49 лет, проживающих на территории России (71 мужчин (31%), 159 женщин (69%)). В исследовании приняли участие жители следующих регионов России: Центральный округ (40,1%), Северо-Западный (10,4%), Уральский (5,1%), Приволжский (27,4%), Южный (10,5%), Сибирский (6,5%).

Участников набирали с помощью объявлений в социальных сетях, включая Facebook и ВКонтакте, или с помощью прямой ссылки, отправленной участникам по электронной почте. Участие в исследовании было добровольно. Все участники были проинформированы о целях исследования и дали информированное согласие на обработку персональных данных.

Методика. Субъектам задавали конкретную информацию, связанную с COVID-19, такую как наличие/отсутствие заболевания и случаев гибели среди родственников, близких друзей или знакомых от COVID-19. В ходе исследования у респондентов диагностировались личностные характеристики: черты личности, стратегии совладания со стрессом, уровень ситуативной (реактивной) тревожности. Для выявления этих индивидуально-психологических характеристик были использованы следующие методики:

– русскоязычная версия методики Big Five Inventory-2 [19]. Данная методика включает 60 вопросов и измеряет пять черт личности: экстраверсию (E), доброжелательность (A), добросовестность (C), нейротизм (N), открытость опыту (O);

– методика «Индикатор копинг-стратегий» [4]. Эта методика позволяла выявить три основные стратегии совладания со стрессом: разрешение проблем, поиск социальной поддержки, избегание проблем;

– методика диагностики ситуативной (реактивной) тревожности Ч.Д. Спилбергера, Ю.Л. Ханина [20].

Также субъекты заполняли опросник CAS для оценки уровня тревожности, связанного с COVID-19. Опросник CAS – это шкала самоотчета, предназначенная для измерения уровня дисфункциональной тревожности, связанной с COVID-19. Он состоит из пяти вопросов, каждый из которых оценивается по пятибалльной шкале Лайкерта (т.е. ответы варьируются от никогда до почти каждый день в течение последних 2 недель). Если общая сумма баллов равна 9 или выше, диагностируется дисфункциональная тревога, связанная с COVID-19 [17, 18].

Статистический анализ. Для изучения полученных данных использовались методы описательной статистики. Для количественных переменных рассчитывались среднее (M) и стандартное отклонение (SD). Для качественных (номинальных) данных использовался частотный анализ (абс., %). Сравнения между группами проводились с использованием t-критерия Стьюдента. Корреляция Спирмена была рассчитана для оценки взаимосвязей между психологическими переменными, социально-демографическими характеристиками и данными опросника CAS. Также использовался критерий Пирсона (критерий χ^2) для сравнения номинальных данных. Анализ проводился с использованием программы Statistica 12.0 для Windows.

Результаты

Социально-демографические и личностные характеристики выборки, а также баллы по шкале CAS обобщены в табл. 1, 2.

Т а б л и ц а 1

Социально-демографическая характеристика выборки

Признак		N (%)	Баллы по CAS M±SD
Пол	Мужчины	71 (31)	2,87±2,64
	Женщины	159 (69)	3,13±2,01
Возраст	<20	62 (27)	2,91±2,13
	20–29	97 (42)	3,07±2,19
	30–39	31 (14)	3,17±2,57
	40–49	40 (17)	2,83±1,97
Уровень образования	Высшее	87 (38)	3,14±2,28
	Незаконченное высшее	108 (47)	3,02±2,55
	Среднее профессиональное	8 (4)	3,88±3,19
	Среднее	27 (11)	2,54±1,94
Болели COVID-19	Нет	163 (71)	2,71±1,55
	Да	67 (29)	3,88±2,56
Заражение COVID-19 среди близких	Нет	35 (15)	3,91±2,08
	Да	195 (85)	3,89±2,84
Смертность среди близких от COVID-19	Нет	161 (70)	2,47±1,14
	Да	69 (30)	4,41±2,54

Примечание. Характеристики выборки описывают количество субъектов (N), процент (%), среднее значение (M) и стандартное отклонение (SD).

Данные личностных характеристик и CAS в общей выборке

Название шкал	Параметр	M±SD
Тест Спирбергера–Ханина	Реактивная тревожность	43,34±12,87
Индикатор копинг-стратегий	Разрешение проблем	26,78±4,64
	Поиск социальной поддержки	22,75±5,63
	Избегание проблем	20,3±5,27
	Экстраверсия (E)	36,91±7,54
Big Five Inventory–2	Доброжелательность (A)	45,11±6,77
	Добросовестность (C)	45,04±7,6
	Нейротизм (N)	35,06±8,56
	Открытость опыту (O)	43,47±7,01
CAS	Уровень тревоги	3,08±1,88

Примечание. Личностные характеристики выборки описывают среднее значение (M) и стандартное отклонение (SD).

Корреляционный анализ между социально-демографическими показателями и CAS представлен в табл. 3. Выявлена статистически значимая положительная корреляция между наличием смертности среди близких от COVID-19 и баллом по шкале CAS ($r = 0,44$).

Таблица 3

Данные корреляционного анализа между социально-демографическими показателями и CAS

Показатель	CAS
Возраст	-0,04
Пол	-0,04
Образование	-0,05
Болели COVID-19	0,04
Заражение COVID-19 среди близких	0,09
Смертность среди близких от COVID-19	0,44*

Примечание. Представлены коэффициенты корреляции Спирмена (r); * – уровень статистической значимости при $p < 0,05$.

Корреляционный анализ между личностными характеристиками и CAS представлен в табл. 4. Выявлена статистически значимая положительная корреляция между CAS и уровнем реактивной тревожности ($r = 0,36$), поиска социальной поддержки ($r = 0,25$), избегания проблем ($r = 0,27$), нейротизма ($r = 0,43$). Также обнаружена обратная корреляция между CAS и экстраверсией ($r = -0,30$), доброжелательностью ($r = -0,27$), добросовестностью ($r = -0,29$).

Таблица 4

Данные корреляционного анализа между личностными характеристиками и CAS

Характеристики	CAS
Реактивная тревожность	0,36*
Разрешение проблем	-0,01
Поиск социальной поддержки	0,25*
Избегание проблем	0,27*
Экстраверсия (E)	-0,30*
Доброжелательность (A)	-0,27*
Добросовестность (C)	-0,29*
Нейротизм (N)	0,43*
Открытость опыту (O)	-0,11

Примечание. Представлены коэффициенты корреляции Спирмена (r); * – уровень статистической значимости при $p < 0,05$.

Что касается опросника CAS, то оценка ≥ 9 указывает на наличие дисфункциональной тревоги, связанной с коронавирусом. В этом исследовании 50 субъектов показали оценку ≥ 9 ,

что составляет примерно 21,74% от общей выборки. На основании этих данных мы дополнительно выделили две группы сравнения: с наличием дисфункциональной тревоги, связанной с COVID-19 ($CAS \geq 9$), и контроль (отсутствие дисфункциональной тревоги) ($CAS < 9$). Сравнение групп по социально-демографическим параметрам также показало, что у лиц с наличием дисфункциональной тревоги, связанной с COVID-19, чаще выявлялись потери (гибель) среди близких в результате заражения COVID-19 по сравнению с контролем ($p = 0,045$) (рис. 1).

Потеряли ли Вы кого-то из близких в результате заражения и/или тяжелых последствий COVID-19?

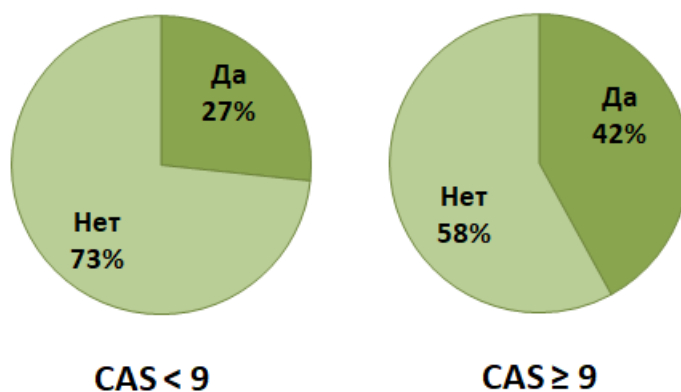


Рис. 1. Смертность среди близких от COVID-19 в исследуемых группах

При сравнении групп по личностным характеристикам у лиц с наличием дисфункциональной тревоги, связанной с COVID-19, обнаружены статистически значимо более высокие показатели реактивной тревожности ($p < 0,001$), поиска социальной поддержки ($p = 0,03$), избегания проблем ($p < 0,001$) и нейротизма (N) ($p < 0,001$), а также меньшие показатели экстраверсии (E) ($p < 0,001$), доброжелательности (A) ($p < 0,001$) и добросовестности (C) ($p < 0,001$) по сравнению с контролем (рис. 2).

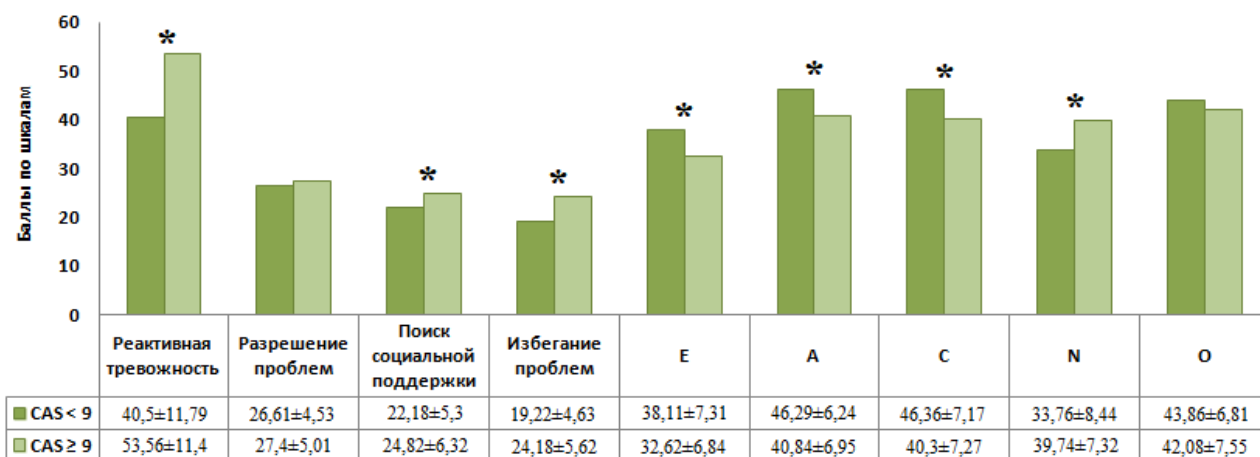


Рис. 2. Личностные характеристики в исследуемых группах

Обсуждение

Целью этого исследования было изучение уровней страха и тревоги, связанного с COVID-19, с помощью Шкалы тревоги по поводу коронавируса (CAS) среди жителей России. В частности, в исследовании изучалось влияние на показатели CAS таких социально-демографических переменных, как пол, возраст, уровень образования и ситуационный опыт,

связанный с COVID-19 (положительный результат на COVID-19, заражение и гибель среди близких от COVID-19). Кроме того, это исследование было направлено на изучение того, могут ли личностные характеристики влиять на уровень страха и беспокойства по поводу коронавируса.

Это исследование наглядно демонстрирует взаимосвязь между рассматриваемыми переменными. Показатели тревоги и страха, измеренные с помощью CAS, имели положительную корреляцию: субъекты с более высокими баллами по CAS, чаще сталкивались с гибелью среди близких в результате заражения и/или тяжелых последствий COVID-19. Аналогичные результаты были обнаружены при анализе групп, с наличием/отсутствием дисфункциональной тревоги в отношении COVID-19. Эти данные согласуются с другими выводами, предполагающими, что, как правило, лица, сталкивающиеся с потерями среди близких в результате какого-либо заболевания, испытывают большие опасения и чувство страха и тревоги в отношении данного заболевания [21].

Кроме того, обнаружено, что на уровень тревоги в отношении COVID-19 влияют стабильные психологические черты личности, такие как реактивная тревожность, поиск социальной поддержки, избегание проблем, экстраверсия, доброжелательность, добросовестность и нейротизм. Полученные данные демонстрируют, что высокие показатели CAS (в том числе наличие дисфункциональной тревоги, связанной с COVID-19) свойственны лицам с более высоким уровнем реактивной тревожности, поиска социальной поддержки, избегания проблем и нейротизма, а также с меньшей экстраверсией, доброжелательностью и добросовестностью. При этом наибольший вклад в уровень тревоги перед коронавирусом оказывал нейротизм, что согласуется с данными исследования J.V. Lippold et al. (2020) [22].

Заключение

Таким образом, в результате проведенного исследования показано, что на уровень тревоги, связанный с COVID-19, существенно оказывает влияние наличие потерь среди близких от коронавируса, а также личностные особенности респондентов, в особенности нейротизм. Полученные данные могут быть полезны при разработке программ, с целью уменьшить страх общества перед коронавирусом.

Список источников

1. Онлайн статистика коронавируса в мире: графики и таблица по странам // URL: www.gogov.ru/covid-19/world (дата обращения: 16 февраля 2022 г.).
2. Pascarella G., Strumia A., Piliago C., Bruno F., Del Buono R., Costa F., Scarlata S., Agrò F.E. COVID-19 diagnosis and management: a comprehensive review // *J Intern Med.* 2020. № 2. Pp. 192–206.
3. Галкин С.А., Пешковская А.Г., Иванова С.А., Бохан Н.А. Когнитивные нарушения у пациентов, перенесших COVID-19 // *Якутский медицинский журнал.* 2021. № 1 (73). С. 69–72.
4. Orrù G., Marzetti F., Conversano C., Vaghegini G., Miccoli M., Ciacchini R., Panait E., Gemignani A. Secondary Traumatic Stress and Burnout in Healthcare Workers during COVID-19 Outbreak // *Int J Environ Res Public Health.* 2021. № 1. P. 337.
5. Peshkovskaya A. Letter to the editor: Other Consequences. COVID-19 and underestimated public health crisis // *Journal of Psychiatric Research.* 2021. № 144. P. 320–322.
6. Di Giuseppe M., Zilcha-Mano S., Prout T.A., Perry J.C., Orrù G., Conversano C. Psychological Impact of Coronavirus Disease 2019 Among Italians During the First Week of Lockdown // *Front Psychiatry.* 2020. № 11. P. 576597.
7. Brooks S.K., Webster R.K., Smith L.E., Woodland L., Wessely S., Greenberg N., Rubin G.J. The psychological impact of quarantine and how to reduce it: rapid review of the evidence // *Lancet.* 2020. № 395. P. 912–920.
8. Germani A., Buratta L., Delvecchio E., Mazzeschi C. Emerging Adults and COVID-19: The Role of Individualism-Collectivism on Perceived Risks and Psychological Maladjustment // *Int J Environ Res Public Health.* 2020. № 10. P. 3497.
9. Lee S.A., Mathis A.A., Jobe M.C., Pappalardo E.A. Clinically significant fear and anxiety of COVID-19: A psychometric examination of the Coronavirus Anxiety Scale // *Psychiatry Res.* 2020. № 290. P. 113112.
10. Lopez A., Caffò A.O., Tinella L., Di Masi M.N., Bosco A. Variations in mindfulness associated with the COVID-19 outbreak: Differential effects on cognitive failures, intrusive thoughts and rumination // *Appl Psychol Health Well Being.* 2021. № 4. P. 761–780.

11. Iachini T. Psychological and situational effects on social distancing and well-being during the COVID-19 pandemic: not a question of real risk // *Front. Psychol.* 2020. № 11. P. 577684.
12. Qian D., Li O. The Relationship between Risk Event Involvement and Risk Perception during the COVID-19 Outbreak in China // *Appl Psychol Health Well Being.* 2020. № 4. P. 983–999.
13. Галкин С.А., Пешковская А.Г., Рощина О.В., Кисель Н.И., Иванова С.А., Бохан Н.А. Особенности мозговой активности при алкогольной зависимости в задаче на ингибиторный контроль // *Бюллетень сибирской медицины.* 2020. № 4. С. 38–45.
14. Воеводин И.В., Пешковская А.Г., Галкин С.А., Белокрылов И.И. Социальная адаптация и психическое здоровье студентов-мигрантов в Сибири // *Социологические исследования.* 2020. № 11. С. 157–161.
15. Torales J., O'Higgins M., Castaldelli-Maia J.M., Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health // *Int J Soc Psychiatry.* 2020. № 4. P. 317–320.
16. Huang Y., Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey // *Psychiatry Res.* 2020. № 288. P. 112954.
17. Lee S. A. Coronavirus Anxiety Scale: A brief mental health screener for COVID-19 related anxiety // *Death Stud.* 2020. № 7. P. 393–401.
18. Mozzoni M., Franzot E. Coronavirus Anxiety Scale (CAS) // Department of Psychology. 2020.
19. Soto C.J., John O.P. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power // *J Pers Soc Psychol.* 2017. № 1. P. 117–143.
20. Spilberger C. Anxiety study. Diagnosis of emotional and moral development // Ed. and comp. Dermanova I.B. 2002. P. 124–126.
21. Menzies R.E., Menzies R.G. Death anxiety in the time of COVID-19: theoretical explanations and clinical implications // *Cogn Behav Therap.* 2020. № 13. P. 9.
22. Lippold J.V., Laske J.I., Hogeterp S.A., Duke É., Grünhage T., Reuter M. The Role of Personality, Political Attitudes and Socio-Demographic Characteristics in Explaining Individual Differences in Fear of Coronavirus: A Comparison Over Time and Across Countries // *Front Psychol.* 2020. № 11. P. 552305.

RESEARCH OF THE LEVEL OF ANXIETY ASSOCIATED WITH COVID-19: THE INFLUENCE OF SOCIO-DEMOGRAPHIC AND PERSONAL CHARACTERISTICS

S.A. Galkin^{1,2}, D.G. Naidenko², V.L. Goiko²

¹ *Tomsk National Research Medical Center of the Russian Academy of Sciences, Tomsk, Russian Federation*

² *National Research Tomsk State University, Tomsk, Russian Federation*

Abstract. The assessment of the level of anxiety associated with COVID-19, as well as the impact of socio-demographic and personal characteristics on it, was carried out. During the online testing, 230 people aged 18 to 49 were interviewed. It was found that 50 (21.74%) respondents had dysfunctional anxiety associated with coronavirus according to the CAS questionnaire. It was also found that the level of anxiety associated with COVID-19 was significantly influenced by mortality from coronavirus among relatives ($r = 0.44$; $p = 0.045$), as well as some personal characteristics of respondents: the level of reactive anxiety ($r = 0.36$; $p < 0.001$), the search for social support ($r = 0.25$; $p = 0.03$), avoidance of problems ($r = 0.27$; $p < 0.001$), extraversion ($r = -0.30$; $p < 0.001$), benevolence ($r = -0.27$; $p < 0.001$), conscientiousness ($r = -0.29$; $p < 0.001$) and neuroticism ($r = 0.43$; $p < 0.001$). The data obtained can be useful in the development of programs to reduce anxiety and public fear of coronavirus.

Keywords: new coronavirus infection, COVID-19, anxiety, fear, CAS, personality traits

МЕТААНАЛИЗ ПОНЯТИЯ «КУЛЬТУРНЫЙ КОД»

Э.В. Гмызина, Т.С. Городилова

Вятский государственный университет, Киров, Россия

Аннотация. В статье описываются результаты метаанализа зарубежных и российских научных публикаций, посвященных культурному коду. Методологическую основу исследования составили методы аналитического обобщения, измерения, статистической обработки полученной информации. На примере сравнительного анализа библиографических баз данных eLIBRARY.RU и Scopus показывается специфика методологических подходов к определению культурного кода и выделению его элементов, структуры. В ходе проведенного исследования была разработана система фильтров, позволяющая определить наиболее релевантные источники. Среди значимых элементов публикаций рассматривались научная дисциплина, в рамках которой проводился анализ концепта «культурный код», базовый методологический подход автора, теоретическая или прикладная направленность работы. Были выявлены основные методологические подходы исследований культурного кода: аксиологический, лингвокультурологический, семиотический, социокультурный, психоаналитический. Исследование показало растущий интерес научного сообщества к проблематике культурного кода и необходимости систематизации имеющихся данных для выработки более эффективных подходов в его изучении.

Ключевые слова: культурный код, метаанализ, систематический обзор, методология гуманитарных исследований.

Феномен культурного кода привлекает интерес исследователей во всем мире, но особенно актуален он для российского научного сообщества. Первые публикации, посвященные культурному коду, появлялись еще в 1980–1990-е гг., а с 2000-х гг. наблюдается их неуклонный количественный рост. Однако попытки продвинуться в формировании общей перспективы подобных исследований все еще наталкиваются на размытость, аморфность самого понятия «культурный код», в осмыслении которого сохраняется множество подходов, точек зрения, нередко конфликтующих между собой. Тот факт, что данный термин в последнее время стал активно употребляться не только учеными, но и, как справедливо отмечает В.М. Савицкий, «философствующими эссеистами, публицистами и тянущимися за ними чиновниками», привел к тому, что произошло «размывание соответствующего понятия» [1, с. 68]. Концепт «культурный код» трактуется исследователями то слишком узко, то чрезмерно широко. Сосредоточение авторов на одной из составляющих культурного кода приводит к погружению в изучение выбранного аспекта и к неразличению прочих ключевых и второстепенных элементов.

Выявление структуры концепта «культурный код» необходимо в качестве методологической основы для изучения кодов определенных культур, для настройки процедур декодирования и интерпретации текстов культур в современном межкультурном диалоге. В этих условиях, обращение к методологии метаанализа позволяет объединить собранные из различных источников результаты исследований.

Метаанализ – это метод статистического анализа, который позволяет дать всестороннее рассмотрение большого количества результатов предыдущих исследований по определенной теме. Один из его разработчиков, Джин В. Гласс, описал метаанализ как «анализ анализов», подчеркнув, что использует его «для обозначения статистического анализа большой коллекции результатов анализа отдельных исследований с целью интеграции результатов» [3, с. 3]. Большинство метааналитических исследований можно рассматривать как продвинутый тип систематического обзора, поскольку, в отличие от других типов обзоров, метаанализ направлен на количественный синтез доказательств. Обычно целью является получение числовой оценки, описывающей четко очерченную область исследований в целом.

Методология метаанализа эффективно работает в рамках естественно-научных исследований, особенно в медицине, в которой дает большую выборку для анализа, хорошую статистическую мощь; может быть использована для повышения доказательности и прогнозирования. Однако в рамках социально-гуманитарного знания применение метаанализа имеет свою специфику, ряд ограничений и требует определенной адаптации. Легче всего это сделать с эмпирически ориентированными исследованиями, в то время как работы теоретического характера предполагают опору не только на объективные количественные показатели, но и на авторскую интерпретацию ученого, которую необходимо принимать во внимание.

Метаанализ понятия «культурный код обеспечивает систематический и всесторонний анализ публикаций по проблематике культурного кода», схожих с ним терминов, позволяет наметить путь для дальнейших исследований. Целью данного метаанализа является выявление материалов, которые исследуют содержание концепта «культурный код», осуществляют терминологическое закрепление данного явления, выделяют структурные компоненты кода.

Источники данных и стратегия поиска

Систематический поиск опубликованных результатов исследований проводился с использованием библиографических баз данных eLIBRARY.RU, Scopus, Web of Science (до мая 2022 г.). Однако в рамках данной статьи будет описан опыт проведения метаанализа на примере доступных для работы в настоящий момент баз данных eLIBRARY.RU и Scopus.

Результаты поиска в базе данных Scopus

Поисковым термином, используемым при отборе документов, был «cultural code» (cultural AND code). Сначала жесткие критерии исключения не использовались, чтобы обеспечить достаточную выборку исследований для применения описанной ниже процедуры отбора материала.

Поэтому сначала поиск ключевого слова «cultural code» в пределах «Название статьи, краткое описание, ключевые слова» дал результат в 8,032 найденных документов (по состоянию на 07.06.2022). Из них на социальные науки пришлось 4 247 источников, искусство и гуманитарные науки – 2 866, медицину – 1 288 (первые три места, остальные – ниже), инженерные науки – 652, психологию – 631 и т.д.

Поэтому было добавлено ограничение по доступности и типу документа (открытый доступ; статья, глава в книге) – это сократило количество материалов до результата в 1 658 источников. Ограничение по отрасли знания (social sciences, arts and humanities) привело к результату в 998 документов, но там все еще было довольно много материалов, в которых понятие кода и культуры используются хоть и взаимосвязано, но отдельно друг от друга и нередко в сочетании с понятием «cultural factor».

Одной из самых первых работ, где встречается в качестве самостоятельного предмета исследования концепт «культурный код» стала работа 1999 г. «Cultural codes – who holds the key?: The concept and conduct of evaluation in central and eastern Europe» британских исследователей J. Nyatt и H. Simons. В данной статье авторы рассматривают влияние культурных кодов на поведение и ценности людей, а сами культурные коды определяются как символы и системы значений, которые имеют отношение к представителям определенной культуры (или субкультуры) [2]. Обращает на себя внимание тот факт, что феномен культурного кода привлек интерес с позиции межкультурного анализа (на примере изучения поведения людей в Центральной и Восточной Европе), который задает аксиологическую, семиотическую направленность, популярную в более поздних исследованиях.

В целом же, удельный рост публикаций по тематике культурного кода наблюдается уже после 2000-х гг., что обусловило дальнейший отбор статей по хронологическому признаку.

Критерии включения. В итоге оптимальными критериями для поисковых запросов стали: ключевое слово, название статьи, аннотация, категории доступа («полный текст», «реферат»), годы издания (2012–2022), отрасль знания (социальные науки, искусство и гуманитарные науки. Языковые ограничения не вводились.

Результаты поиска. После применения описанных выше фильтров осталось 23 документа (21 статья – 91,3 %, 2 документа (8,7 %) – глава в книге). Самое большое количество в году (6 статей) пришлось на 2020 г., по 3 статьи было опубликовано в 2015 и 2018 гг.

Критерии исключения. Далее был проведен иерархический отбор литературы путем анализа заголовков статей и исключения явно нерелевантных публикаций, в которых культурный код просто упоминается просто как словосочетание или «размытое понятие».

Дальнейшее разграничение статей путем более глубокого изучения аннотаций и полного текста (у 10 публикаций он стал недоступен для российских исследователей, но определенные выводы можно сделать по тексту реферата) позволило сузить объем исследований в соответствующей предметной области. Все потенциально подходящие статьи были извлечены и изучены без предвзятости в отношении их выводов.

Выводы по результатам поиска. Распределение документов по странам показало, что наибольшее число публикаций (7 статей) приходится на долю российских авторов (сейчас доступ из БД Scopus к российским текстам также невозможен). Второе место занимает Великобритания (3 статьи), далее по 2 статьи опубликованы в Чехии, Израиле, Польше, Испании, и по одной статье приходится на Францию, Грецию, Италию и Казахстан. Очевидно, что эти цифры демонстрируют заинтересованность в изучении феномена культурного кода преимущественно российским научным сообществом. Все авторы представляют рассмотрение тематики культурного кода только в одной статье, поэтому какой-то явной закономерности в «лидерах» исследований и принадлежности к научной школе не наблюдается.

Распределение публикаций по отраслям знания получилось следующим: социальные науки – 54,3% источников, искусство и гуманитарные науки – 42,9%, психология – 2,9%. В интерпретации понятия культурный код преобладают подходы, которые условно можно назвать лингвокультурологическим и социокультурным (примерно по 35% работ в рамках каждого), достаточно популярны семиотический, аксиологический и психоаналитический подходы. В тоже время, часть из этих работ к трактовке культурного кода подходят с позиций междисциплинарности, работая на «стыке наук»: лингвистики, педагогики, филологии, антропологии.

В первую тройку по количеству цитирований вошли следующие статьи: D. Gavriely-Nuri «Cultural approach to CDA» (Израиль, 2012 г., 20 цитирований), H. Rodriguez-Navarro, A. García-Monge, M.C. Rubio-Campos «The process of integration of newcomers at school: Students and gender networking during school recess» (Испания, 2014 г., 11 цитирований), G. Yair «Hierarchy versus symmetry in German and Israeli science» (Израиль, 2020 г., 5 цитирований).

Анализ работ зарубежных исследователей показывает, что основное внимание уделяется изучению культурного кода в его взаимодействии с языком, системой ценностей (возрастных, этнических, мировоззренческих и т.д.), связан с проблемами национального образования и идентичности.

Однако, в отличие от материалов, представленных в российских изданиях, нет ни одной работы за указанный период, посвященной трактовке самого понятия культурный код. Чаще всего, культурный код выступает как один компонент анализа или фактор, влияющий на те проблемы, которые анализируются в качестве предмета исследования.

Результаты поиска в базе данных eLIBRARY.RU

Критерии включения. Поиск ключевого слова «культурный код» в пределах «название статьи, аннотация, ключевые слова, полный текст публикации» дал результат в 1 497 документов (по состоянию на 07.06.2022). Из них в РИНЦ входят 1 408 публикаций, в ядро

РИНЦ – 93 публикации. Кроме того, были обнаружены документы, с близкими по значению ключевыми словами «национально-культурный код» (31), «национальный культурный код» (18), «культурный код нации» (17), «русский культурный код» (10), «социально-культурный код» (5). Из них на языковедение пришлось 629 источников, литературоведение и фольклористику – 142, культурологию – 118, философию – 118, общественные науки в целом – 106, образование и педагогику – 82, историю – 76, социологию – 40, искусствоведение – 39, экономические науки – 37, политические науки – 22, строительство и архитектуру – 14, право и юридические науки – 13, массовые коммуникации – 9. Единичные публикации отмечены в таких областях знания, как физическая культура и спорт, география, здравоохранение, религия, биология, демография, легкая промышленность, физика и химия.

Первая публикация по ключевому слову «культурный код» в базе eLIBRARY.RU относится к 2003 г. С 2003 по 2006 г. поступило 5 документов. Стремительный рост отмечается с 2008 г.: в 2008 г. – 11, 2013 г. – 61, 2018 г. – 168, 2021 г. – 246.

Распределение документов по типам публикаций показало, что наибольшее количество публикаций приходится на долю научных статей в журналах (904), затем идут статьи в сборниках трудов конференций (452), статьи в сборниках статей (62), статьи в журналах по материалам конференций (22), тезисы докладов на конференциях (21), главы в книгах (10), монографии (5), сборники статей (2). Единично представлены сборник трудов конференции, учебное пособие и автореферат диссертации.

Поиск по ключевым словам свидетельствует, что с понятием «культурный код» в выделенных публикациях чаще всего связаны термины «культура» (144), «концепт» (66), «лингвокультурология» (58), «язык» (58), «языковая картина мира» (55), «символ» (51), «картина мира» (46), «семиотика» (37), «менталитет» (30), «фразеологизм» (29), «художественный текст» (29), «текст» (28), «межкультурная коммуникация» (27), «метафора» (27), «глобализация» (26), «интертекстуальность» (26), «лингвокультура» (26), «образ» (26), «коммуникация» (25), «традиции» (25), «знак» (24), «идентичность» (24).

В отличие от Scopus в российской базе eLIBRARY.RU можно выделить «лидеров» исследований и их принадлежность к определенным научным школам.

В первую тройку по количеству публикаций вошли сотрудник факультета филологии и национальной культуры Рязанского государственного университета им. С. А. Есенина, кандидат культурологии, доцент Н.Г. Меркулова (15 публикаций за период с 2014 по 2021 г.), сотрудник Чувашского государственного университета им. И. Н. Ульянова, доктор филологических наук, профессор Л.В. Борисова (10 публикаций с 2016 по 2022 г.), сотрудник Санкт-Петербургского государственного университета промышленных технологий и дизайна, кандидат технических наук, доцент В.Л. Жуков (10 публикаций с 2017 по 2019 г.).

Лидерами по числу цитирований стали следующие статьи: В.Л. Тамбовцев «Миф о культурном коде в экономических исследованиях» (журнал «Вопросы экономики», 2015 г., 55 цитирований), Д. Александер, Ф. Смит, С. Джакупова «Сильная программа в культурсоциологии» (журнал «Социологическое обозрение», 2010 г., 54 цитирования), Г.Г. Меркулова «Менталитет – культурный код – язык культуры: к вопросу о корреляции понятий» (журнал «Регионоведение», 2015 г., 27 цитирований).

Статистический анализ документов показал, что сформировалось несколько отечественных научных центров по исследованию культурного кода. Среди них Московский государственный университет им. М.В. Ломоносова (37 публикаций), Башкирский государственный университет (28 публикаций), Санкт-Петербургский государственный университет (26 публикаций).

Критерии исключения. С целью выявления публикаций, позволяющих рассмотреть методологические подходы к изучению концепта «культурный код», были применены дополнительные фильтры. Во-первых, из списка публикаций были исключены работы, не соответствующие социально-гуманитарной направленности; во-вторых, исключены работы, направленные на изучение конкретного эмпирического материала и не претендующие на

осмысление теоретических и методологических аспектов «культурного кода»; в-третьих, были исключены статьи, в которых отсутствует список ключевых слов, аннотация, ссылки на цитируемые источники. В корпус текстов также не попали редакционные статьи, рецензии на различные научные издания.

Результаты поиска. В результате применения фильтров сформировалась выборка из 658 документов, опубликованных с 2003 по 2022 г. Наибольшее количество статей пришлось на журналы «Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика», «Международный журнал исследований культуры», «Вопросы когнитивной лингвистики», «Сибирский филологический журнал», «Филологические науки. Научные доклады высшей школы», «Филологические науки. Вопросы теории и практики», «Культура и цивилизация», «Казанская наука», «Общество: философия, история, культура», «Когнитивные исследования языка».

Выводы по результатам поиска. С помощью контент-анализа определены основные исследовательские методы, частотность ключевых слов, основные направления научной мысли в выявлении структурных компонентов культурного кода. В процессе анализа фиксировались такие характеристики публикаций, как а) дисциплина, в рамках которой анализируется концепт «культурный код»; б) объект исследования; в) преобладающий в публикации исследовательский подход. Результаты анализа показали, что в ряде случаев темы публикаций носили междисциплинарный характер, находились на стыке двух и более дисциплин, но в целом можно утверждать, что изучением культурного кода преимущественно занимаются лингвисты, культурологи, литературоведы, фольклористы, философы, социологи, журналисты, психологи.

Анализ работ отечественных исследований показывает, что наибольшее внимание уделяется изучению культурного кода в таких контекстах, как язык, символика, менталитет, традиции, ценности, идентичность.

В результате проведенного анализа были выявлены ведущие методологические подходы в работах отечественных исследователей. Они обозначены исходя из сложившейся в социально-гуманитарной сфере исследовательской практики следующим образом: аксиологический подход, лингвокультурологический подход, семиотический подход, социокультурный подход, психоаналитический подход. В большинстве статей, попавших в выборку, преобладал один ведущий методологический подход. Анализ показал следующие результаты: семиотический подход – 229 статей (34,8%), социокультурный подход – 225 статей (34,2%), лингвокультурологический подход – 133 статьи (20,2%), аксиологический подход – 62 статьи (9,4%), психоаналитический подход – 9 статей (1,4%).

Таким образом, метаанализ, объединяя статистические и логические процедуры анализа и обработки данных, позволяет сформировать достаточно точную, объективную картину научных исследований, посвященных культурному коду. Результаты проведенного метааналитического исследования концепта «культурный код» показывают, что несмотря на растущий интерес данный феномен все еще недостаточно осмыслен в теоретико-методологическом аспекте. В то же время на основе разрозненных исследований, которые публикуются в отечественных и зарубежных изданиях можно говорить о постепенном формировании теоретической основы для изучения культурного кода. В потоке научной информации о культурном коде можно отметить явное преобладание авторов из России, Белоруссии, Украины, Казахстана, Китая, Индии. В целом, теория культурного кода на Евразийском пространстве развивается намного динамичнее, чем в странах Запада. Этот вывод, без сомнения, следует подкрепить дополнительными компаративными исследованиями, но обозначенная тенденция просматривается достаточно четко.

Список источников

1. Савицкий В.М. Культурные коды: сущность, состав и функционирование в процессе общения // Дискурс профессиональной коммуникации. 2019. Т. 1, № 4. С. 68–77.

2. Hyatt, J., Simons, H. Cultural codes – who holds the key?: The concept and conduct of evaluation in central and eastern Europe // *Evaluation*. 1999. Vol. 5 (1). P. 23–41
3. Glass Gene V. Primary, Secondary, and Meta-Analysis of Research // *Educational Researcher*. 1976. Vol. 5 (10). P. 3–8.

META-ANALYSIS OF THE CONCEPT OF CULTURAL CODE

E.V. Gmyzina, T.S. Gorodilova

Vyatka State University, Kirov, Russian Federation

Abstract. The article describes the results of a meta-analysis of foreign and Russian scientific publications devoted to the cultural code. The methodological basis of the study was the methods of analytical generalization, measurement, statistical processing of the information received. On the example of a comparative analysis of bibliographic databases eLIBRARY.RU and Scopus, the authors demonstrate the specificity of methodological approaches to the definition of the cultural code and the allocation of its elements and structure. In the course of the conducted research, a filter system was developed to identify the most relevant sources. Among the significant elements of the publications were the subject field within which the concept of cultural code was analyzed, the author's basic methodological approach, the theoretical or applied nature of the work. The main methodological approaches to the research of the cultural code were identified: axiological, linguoculturological, semiotic, sociocultural, psychoanalytic. The study showed the growing interest of the scientific community in the problems of the cultural code and the need to systematize the available data to develop more effective approaches to its study.

Keywords: cultural code, meta-analysis, systematic review, methodology of humanitarian research

РОСДИСТАНТ КАК ИННОВАЦИОННЫЙ КОНЦЕПТ НОВОГО ПОКОЛЕНИЯ ОБУЧЕНИЯ СТУДЕНТОВ ВУЗОВ В ИНФОРМАЦИОННОМ ПРОСТРАНСТВЕ

Т.Н. Иванова, А.С. Новичкова

Тольяттинский государственный университет, Тольятти, Россия

Аннотация. В информационном обществе структура творческих производительных сил человека актуализируется в форме человеческих информационных ресурсов. Человек представляется как главная движущая сила общества, он играет разнообразные общественные роли в сложной структурной организации. В силу развития и информатизации общества в образовании начинают внедрять систему дистанционного обучения. Характерным для дистанционного обучения будет то, что обучаемый сам получает знания путем изучения информационных ресурсов. Дистанционное обучение позволяет использовать различные визуальные материалы для более интенсивной и интересной работы. В связи с сегодняшней обстановкой в мире, распространение вируса COVID-19 подтолкнуло большинство социальных институтов перейти на удаленную работу. Дистанционное образование стало актуальным форматом для обучения студентов.

Ключевые слова: дистанционное обучение, информационное пространство, университет, росдистант, концепт, инновации, новое поколение

Во второй половине XX века начинают стремительно развиваться информационные технологии, и вследствие увеличивается объем накопленной информации. Для современного мира стало нормой постоянное увеличение объемов информации, что послужило началу постановки задачи с развитием единого информационного пространства [1].

В рамках реализации «Стратегии развития информационного общества в Российской Федерации на 2017–2030 года» приняты цели и задачи для успешного развития информационного общества, а также формирования информационного пространства для удовлетворения потребностей общества в получении качественных и достоверных сведений. Определим, что такое информационное пространство. «Информационное пространство – совокупность информационных ресурсов, созданных субъектами информационной сферы, средств взаимодействия таких субъектов, их информационных систем и необходимой информационной инфраструктуры» [2]. Целями в данном пункте будут являться создание такого пространства для постоянного и качественного получения новых знаний и расширения кругозора безопасной информацией. Это все возможно при реализации развития образовательных проектов и создании общедоступной безопасной информационной среды. В рамки выполнения этих целей также будут входить усовершенствование механизмов обмена знаниями, использование дистанционных технологий при реализации образовательных программ, создание площадок для самореализации представителей образовательных организаций.

Опираясь на «Стратегию развития информационного общества РФ на 2017–2030 годы», можно отметить, что вопрос развития дистанционного обучения входит в цели по успешному становлению информационного пространства, поэтому изучение этой темы актуально на сегодняшний день.

Распространение вируса COVID-19 подтолкнуло большинство учебных заведений перейти на удаленное обучение с использованием дистанционной формы организации учебного процесса в массовом порядке. Это дает нам повод глубже изучить работу дистанционного обучения в высших учебных заведениях.

Развитие современной цивилизации характеризует XXI век как эпоху информационного становления, в которой все больше начинают распространяться и совершенствоваться информационные технологии, формируется инфосфера. Также информационное общество характеризуется глобализацией общественных процессов. Организационно-технологической основой информационного общества являются глобальные информационные сети.

Актуальность темы изучения информационного общества мы можем наблюдать и сегодня. С конца 60-х годов XX века сформулировано множество определений информационного общества. Большинство исследователей подразумевают знания как стратегический ресурс. А это значит, одним из базовых принципов определения является уровень доступности для общества информации и технологий. Социолог Даниел Белл пишет: «В наступающем столетии решающее значение для экономической и социальной жизни, для способов производства знания, а также для характера трудовой деятельности человека приобретает становление нового социального уклада, основывающегося на телекоммуникациях» [3].

Известным теоретиком в изучении становления и развития информационного общества считается В. Л. Иноземцев. Наряду с терминами *постиндустриализм*, *информационное общество* использует термин *постэкономическое общество*. Главным отличием от предыдущих терминов выступает значение и роль личности в социальной культуре. Он выдвигает концепцию, в которой законы технологического развития не являются базисными, потому что они не определяют закономерности эволюции социальных отношений. По мнению Иноземцева, мы переходим к социуму, в котором преобладают личностные и психологические связи [4].

В информационном обществе структура творческих производительных сил человека актуализируется в форме человеческих информационных ресурсов. Человек представляется как главная движущая сила общества, он играет разнообразные общественные роли в сложной структурной организации.

Для лучшего применения информационных технологий и раскрытия потенциала различных сфер общества нужно проведение соответствующих исследований, касающихся, состояния тенденций и перспектив развития общества. Для этого понадобятся измерение структуры информационного общества, выявление тенденция влияния информационных технологий на структуру занятости населения, степень воздействия информационных технологий на качество услуг здравоохранения и образования, вычленение влияния на личностные и профессиональные качества человека, постоянное прогнозирование развитие информационного общества.

В книге «Информационное общество как постиндустриальное общество», написанной японским ученым Е. Масудой, описывается еще одна концепция информационного общества. Компьютерные технологии станут фундаментом нового общества путем замещения и усиления умственного труда человека. Такая информационная революция сделает возможным массовое производство когнитивной и систематизированной информации. Также ведущей отраслью экономики будет интеллектуальное производство, продукт которого будет распространяться благодаря новым технологиям. Масуда пишет, что это общество будет бесконфликтным, и выдвигает отличие от индустриального общества: характерной ценностью будет время, а не потребление товаров. Будут возрастая решение насущных проблем и развитие сотрудничества [5].

«Изучая информационное общество можно выделить такие методологические категории:

1. Социальные отношения в информационном обществе.
2. Социальные общности, возникающие в информационно-коммуникативном пространстве.
3. Социальные процессы в информационном обществе: глобализация социальных процессов, информационные войны, компьютерная преступность» [6].

Также мы можем заметить, что с возникновением нового этапа информационного общества у людей стало больше возможностей выражать свое мнение на различные темы в интернете с помощью электронных форумов. Это могут быть обычные бытовые проблемы, но также позиции насчет мировоззренческих идей. Информацию можно не только получать, но отдавать. Процесс глобализации упростил этот процесс, начали возникать новые возможности коммуникации между людьми не только на локальном уровне, но и на международном.

Введение новых запросов к системе образования считается одним из главных приоритетов общества. К совершенствованию системы образования стараются все больше подкреплять развитие критического мышления и навыки soft skills. И для развития профессиональных компетенций востребованным считается применение интерактивных форм обучения.

Сегодня можно выделить следующие актуальные проблемы в образовании:

1) Утрачивается доверие к информационным источникам. Проблема увеличения количества источников приводит к тому, что невозможно определить надежность информации. А это уже приводит к упрощению смысла информации и ее обесцениванию. У. Бек утверждает: «Увеличение научного знания обнаруживает риски как для самой науки и образования, так и для общества в целом, поскольку делает недостижимым объективное знание в силу именного огромного количества информации» [7].

2) Происходит обесценивание классического обучения, основанного на запоминании информации, доступных методах ее передачи, движении от простого к сложному с использованием структурированного материала. Поддаются критике традиционные методы обработки информации.

3) Доступность образования провоцирует огромное количество выпускников. От переизбытка людей, получивших высшее образование, происходит нехватка рабочих мест.

Многие университеты применяют информационные технологии для создания учебных материалов, при проведении исследований, для решения управленческих задач. Использование университетами системы онлайн-обучения также довольно распространено.

Для осуществления дистанционного обучения необходимы телекоммуникационные средства для проведения электронных занятий и подключение к интернет-сети. Также для реализации учебного процесса подключаются такие средства: электронная почта, видеоконференции, чат. Кроме того, должны быть разработаны специальные программные продукты, позволяющие эффективно решать организацию проведения занятий. Такие дистанционные образовательные технологии должны обеспечить обучаемым легкий доступ ко всей нужной информации [8].

Рассмотрим характерные черты дистанционного обучения:

– гибкость – образовательный процесс подстраивается под обучаемого, он может учиться в удобном для себя темпе;

– параллельность – обучение может проводиться при совмещении основной и профессиональной деятельности. Данный критерий является одной из основных причин, почему выбирают дистанционное обучение;

– дальное действие – эффективное обучение может реализовываться на любом расстоянии от образовательного учреждения, что помогает студентам, проживающим далеко от вуза и не имеющим финансов для переезда;

– асинхронность – преподавателю и обучающемуся не нужно работать в одно время, процесс обучения будет проходить по удобному для каждого расписанию [9].

В силу развития и информатизации общества в образование начинают внедрять систему дистанционного обучения. Характерным для дистанционного обучения будет то, что обучаемый сам получает знания путем изучения информационных ресурсов. Дистанционное обучение позволяет использовать различные визуальные материалы для более интенсивной и интересной работы.

Распространение вируса COVID-19 подтолкнуло образовательные учреждения перейти на дистанционный вид обучения. Переход был быстрый, и не все ресурсы были готовы к качественной работе.

Наше внимание привлекла система организации дистанционного образования в Тольяттинском государственном университете через платформу «Росдистант».

В социологическом исследовании кафедры «Социология» Тольяттинского государственного университета приняли участие студенты, оценивающие условия работы дистанционного обучения в ТГУ. Методом анкетирования были опрошены студенты ТГУ, дифферен-

цированные по форме обучения, обучающиеся на платформе «Росдистант», которые были переведены на дистанционное обучение из-за эпидемиологической ситуации и впоследствии учились по смешанной форме обучения. Под смешанным обучением подразумевалось, что лекции у студентов проходят онлайн, а практики проводятся очно в университете.

Сбор данных был произведен через сервис по созданию опросов Webanketa. Обработка данных была произведена в программе SPSS.

Рассмотрены социально-демографические характеристики участников анкетирования. В опросе приняли участие 65 % женщин и 35 % мужчин. Респонденты, обучающиеся с первого по пятый курс, распределялись так: первый курс – 26%, второй курс – 24%, третий курс – 27%, четвертый курс – 15% и пятый курс – 8%. По направлению обучения респонденты разделились на 60 % студентов гуманитарных направлений и 40% студентов технических специальностей.

Для начала респондентам был задан вопрос о том, как часто они посещают платформу дистанционного обучения, в целях проанализировать временной интервал посещения студентами платформы дистанционного обучения. Варианты ответа были такие: «5–6 раз в неделю», «3–4 раза в неделю», «1–2 раза в неделю», «несколько раз в месяц».

По результатам 36 % респондентов посещали платформу дистанционного обучения 5–6 раз в неделю, 32% респондентов посещали 3–4 раза в неделю. Вариант «1–2 раза в неделю» выбрали 22 % студентов. Несколько раз в месяц на платформу заходило 8 % респондентов. Отметим, что по результатам студенты в целом часто обучаются через платформу (62%), что подразумевает, что студенты заинтересованы в обучении.

Далее был задан вопрос о том, как проходила сдача домашнего задания на платформе «Росдистант». Варианты были предложены такие: «Все задания были выполнены в срок», «Больше половины заданий были выполнены в срок», «Меньше половины заданий были выполнены в срок», «Все задания были не выполнены». Ответы на данный вопрос распределены далее. Выполнили все задания в срок 58% респондентов, больше половины заданий выполнили 32 % респондентов, и меньше половины заданий были выполнены в срок у 10% респондентов. Отметим, что больше половины респондентов (90%) сдают большинство заданий в срок, что показывает успешную плановость протекания учебного процесса.

Для нас было важно узнать, сколько времени в день у студентов занимает выполнение домашнего задания и изучение учебных материалов. По форме обучения результаты распределились так, что студенты смешанного обучения и студенты «Росдистанта» делали домашнее задание и изучали материал от 1 часа до 5 часов, 72% и 70% ответов соответственно. Более 5 часов в день изучали домашний материал 13% студентов смешанного обучения и 9% студентов «Росдистанта». Тем самым наша гипотеза о том, что у студентов смешанного обучения времени на самостоятельную работу занимало больше, чем у студентов «Росдистанта», не подтвердилась, так как большинство студентов выделяет для изучения материала до 5 часов и формат обучения на это не влияет.

Далее был вопрос о том, какие формы работ применялись у студентов в дистанционном обучении, речь идет о разнообразии форм подачи материалов. У 85% студентов реализуется онлайн-тестирование. Выдача заданий для самостоятельного выполнения производится у 69% респондентов. Показ обучающих презентаций проводится у 56% респондентов. Видеоконференции проходят у 5 % респондентов. Практические занятия проходили у 49% респондентов. Организация переписки с преподавателями через почту была у 39% студентов. Показ видео по теме было у 36% студентов. Доступ к письменным лекциям производился у 36% респондентов.

Чтобы проверить гипотезу о том, что у студентов «Росдистанта» проводится больше форм подачи материала, рассмотрим варианты, где частота была значительно больше на какой-либо форме обучения. На 43% чаще выбирали вариант «проведение видеоконференций» студенты смешанного обучения (77%), чем студенты «Росдистанта» (34%). «Показ обучающих презентаций» на 40% чаще выбирали студенты смешанного обучения

(74%), чем студенты «Росдистанта» (39%). На 20% чаще выбирали вариант «показ видео по теме» студенты смешанного обучения (47%), чем студенты «Росдистанта» (27%).

Таким образом, анализируя результаты исследования, можно сказать, что чаще разнообразные формы подачи материалов встречались у студентов смешанного обучения. Но все виды подачи в большей или меньшей пропорции применялись у студентов «Росдистанта» и студентов смешанного обучения. Так мы можем сказать, что наша гипотеза о том, что у студентов «Росдистанта» применяется больше видов подачи материалов, чем у студентов смешанного обучения, не подтвердилась.

А рассматривая частоты у вариантов с визуальными формами подачи материала, мы можем сказать, что у студентов «Росдистанта» они встречаются реже, чем у студентов смешанного обучения. Это не подтверждает нашу гипотезу о том, что у студентов «Росдистанта» больше ресурсов с визуализацией материала.

Интересно было спросить у студентов, с какими трудностями они столкнулись при обучении в дистанционном режиме. Не понимали, как делать домашнее задание 47% респондентов. Отмечали, что у них были технические трудности 41% респондентов. Материал был сложный для самостоятельного обучения у 38% студентов. У 31% студента была сложность в общении с преподавателем. Плохая скорость интернета была у 15% респондентов. Не было подходящего оборудования для обучения дистанционно у 9% респондентов. Не было трудностей у 19% респондентов.

В дальнейшем мы поинтересовались у опрошенных о том, интересно ли им было обучаться с помощью дистанционных форм подачи материала. Вариант «да» выбрали 36 % респондентов. «Скорее да, чем нет» отметили 37% студентов. Вариант «скорее нет, чем да» выбрали 13% опрошенных. И вариант «нет» отметили 12% студентов.

Интересны будут результаты, распределенные по форме обучения. Мы проанализировали результаты положительных и отрицательных оценок для получения полной картины. Чаще отвечали положительно на вопрос об интересе обучения дистанционно студенты «Росдистанта» (83%), чем студенты смешанного обучения (62%). Отрицательно на этот вопрос на 18% чаще отвечали студенты смешанного обучения (34%), чем студенты «Росдистанта».

Студенты «Росдистанта» положительно отзываются о работе на дистанционной платформе, потому что данную форму обучения они сами выбрали. Но в целом большинство студентов положительно относятся к дистанционному обучению.

Мы также поинтересовались у респондентов о желании продолжить учиться с помощью дистанционных технологий. Больше положительных ответов прослеживается у тех, кто сознательно поступал на дистанционное обучение: 73%, против 55% тех, кто обучается в смешанном режиме.

Таким образом даже если студенты положительно отзываются о проведении дистанционного обучения, это не отменяет наличия проблем, связанных с проведением эффективной работы. Открывая возможности самостоятельного обучения через дистанционную платформу, стоит рассмотреть вопрос, а какие знания получит студент по итогу обучения. Конечно повышаются навыки работы тайм-менеджмента и умения пользоваться информацией, но при этом страдают навыки коммуникации. Упор в дистанционной учебе делается на самостоятельную работу студентов, так как время, проведенное с преподавателями, ограничено. Платформа, на которой происходит обучение, должна компенсировать нехватку живого общения. Также она должна быть современной в функциональном плане, удобной для быстрого отклика на решение задачи, адаптированной для нахождения большого количества пользователей в видеоконференции. Оценка студентами условий учебной площадки будет значимым ресурсом для продвижения системы.

Список источников

1. Васильева М.М. Становление информационного общества в России в условиях глобального информационного пространства // Вестник Московского государственного лингвистического университета. Общественные науки. 2020. № 3 (840). С. 50–58.

2. Указ Президента РФ от 9 мая 2017 г. № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы». URL: <https://base.garant.ru/71670570/> (дата обращения: 13.06.2022).
3. Науменко Т.В. Методологический анализ концепции информационного общества // *Фундаментальные исследования в сфере развития информационного общества*. 2018. № 2. С. 4.
4. Пузикова В.С. Проблема целостного восприятия мира в контексте реалий информационного бытия современного человека // *Современные проблемы массовой коммуникации*. 2014. № 1. С. 3902.
5. Sologub T.G. The Direction of Transformation of Information and Communication Technology (Ict) at the Present Stage of Development into an Electronic and Information Society // *Social Sciences & Humanities*. 2017. № 25. P. 45–58.
6. Лукина Н.П. Информационное общество: состояние и перспективы социально-философского исследования // *Открытое и дистанционное образование*. 2003. № 1. С. 2–6.
7. Рудченко Т.И. Проблемы и перспективы высшего образования в информационном обществе // *Гуманитарные аспекты высшего профессионального образования*. 2018. № 1. С. 134–138.
8. Антонова Н.Л. Особенности дистанционного обучения в вузах России // *Научный электронный журнал Меридиан*. 2020. № 6 (40). С. 168–170.
9. Брежнева С.Н. Плюсы и минусы дистанционного обучения. Из опыта работы. URL: http://www.rusnauka.com/15_NNM_2012/Pedagogica/1_93151.doc.htm (дата обращения: 10.05.2022).

ROSDISTANT AS AN INNOVATIVE CONCEPT OF A NEW GENERATION OF UNIVERSITY STUDENTS' EDUCATION IN THE INFORMATION SPACE

T.N. Ivanova, A.S. Novichkova

Togliatti State University, Togliatti, Russian Federation

Abstract. In the information society, the structure of a person's creative productive forces is actualized in the form of human information resources. A person is presented as the main driving force of society; they play a variety of social roles in a complex structural organization. Due to the development and informatization of society, a distance learning system is being introduced into education. It is characteristic for distance learning that students independently receive knowledge by studying information resources. Distance learning allows one to use various visual materials for more intensive and interesting work. Due to the current situation in the world, the spread of COVID-19 has prompted most social institutions to switch to remote work. Distance education has become a relevant format for teaching students.

Keywords: distance learning, information space, university, Rosdistant, concept, innovations, new generation

ВОЗМОЖНОСТИ ОЦЕНКИ КРЕАТИВНОСТИ ПО ЦИФРОВЫМ СЛЕДАМ

И.А. Исакова

*Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского,
Нижний Новгород, Россия*

Аннотация. В статье представлены результаты исследования, посвященного анализу зарубежного опыта выявления черт характера по цифровым следам в социальных сетях. Изучение проводилось в форме контент-анализа статей из рейтинговых зарубежных журналов, посвященных оценке качеств личности по цифровым следам. Выборка составила 50 статей с высокими показателями цитируемости из библиографической, реферативной базы Scopus за период с 2001 г. по 2021 г. Под цифровыми следами понималась совокупность данных, которые человек осознанно и не осознанно оставляет при использовании Интернета. В результате охарактеризованы положительные и отрицательные стороны использования цифровых следов для составления социально-психологического профиля пользователя; описаны методики оценки черт личности по цифровым следам. В статье также представлен опыт исследования креативности с использованием Big Data лаборатории «Дизайн информации» Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского.

Ключевые слова: цифровизация, креативность, цифровой след, контент-анализ, Интернет, личностные качества, пользователь

Источник финансирования: материал подготовлен в рамках реализации программы стратегического академического лидерства Приоритет 2030.

В XXI в. общество столкнулось с новым глобальным вызовом, которым стал беспрецедентный скачок развития коммуникационных технологий, оказавший значительное влияние на все аспекты его жизнедеятельности. Интернет стал больше, чем просто средством коммуникации, это социальное пространство, формирующее условия взаимодействия нового типа.

Динамика «цифровых» изменений поставила перед социологами новые задачи, связанные с исследованиями трансформировавшихся социальных практик. Однако с первых исследований активности пользователей в глобальном виртуальном пространстве стало понятно, что новые практики общения не всегда можно оценить классическими, формализованными социологическими методиками. Необходимо разрабатывать новые средства онлайн мониторинга, позволяющие осуществлять автоматизированный сбор эмпирического материала, использования математического инструментария для оценки виртуального социального взаимодействия.

Исследования, базирующиеся на цифровых технологиях и использующие возможности сети Интернет, в социологии проводятся уже более 20 лет, а сфера их применения, тематика, методическое разнообразие огромны. Обзор статей в журнале «Мониторинг общественного мнения: экономические и социальные перемены» с 2000 по 2022 г. показал, что тема «технологизации социологии» [1] поднимается с середины первого десятилетия XXI в. Одним из первых этапов осмысления технических возможностей сети Интернет для социологических исследований стало сравнение результатов опросов, проведенных классическими социологическими методами и онлайн. Например, социологи А.В. Шашкин и О.Е. Кузина пришли к выводу, что данные репрезентативны только по отдельным узким темам: «несмотря на то, что Интернет опросы не репрезентативны по отношению к населению России в целом, их можно использовать для решения модельных задач, когда требуется оценить влияние различных факторов на финансовое поведение населения, а также в тех случаях, когда речь идет о труднодоступных для оффлайновых методов пользователях финансовых услуг» [2].

Следующим трендом в развитии эмпирической социологии стал поиск средств онлайн мониторинга социальной активности пользователей. На первых этапах ученые стали приспосабливать средства статистики поисковых систем для исследования социального поведения.

Например, Ф.Н. Ильясов осуществил попытку оценить отношение россиян к теме «финансового кризиса», используя анализ частоты запросов в поисковой системе Яндекс [3].

Со временем стали появляться комплексные методики Интернет-мониторинга. Например, А.А. Азаров и его соавторы предложили методику киберметрического анализа цифровых следов активности пользователей в социальных медиа, совмещающую «ивент-анализ, типологизацию сообщений о рассматриваемых событиях с применением онлайн Интернет-сервиса мониторинга и дискурс-анализ» [4].

На следующем этапе начинает разворачиваться активная дискуссия о преимуществах и недостатках исследования Big Data, которая ведется и в настоящее время.

Обсуждение из теоретического поля, перешло в область решения конкретных исследовательских кейсов. О.Ю. Колесниченко и его соавторы доказывают, что «датафицированная текстовая характеристика «мобильный телефон» в привязке к 2011 г., когда началась «арабская весна», оказалась связана с социально-демократическими процессами в глобальном обществе» [5]. В то же время В.И. Дудина, Д.И. Юдина, на основе сравнения анализа комментариев пользователей видеохостинга с результатами опроса общественного мнения по поводу отношения к фильму, показали, что данные не сошлись [6].

Различные кейсы исследований, проведенных по цифровым следам представлены в работах М.Г. Мягкова и его соавторов (субъективная оценка (не)благополучия населения регионов РФ на основе данных социальных сетей) [7], С.С. Бразевич и коллектива (исследование интересов молодежи через тематический репертуар общения в студенческой среде) [8]. Однако все авторы признают, что работа с Интернет-данными сопряжена с определенным риском. Следует учитывать, что современная глобальная сеть наполнена средствами имитации пользовательской активности, такими как программы «накрутки» просмотров, лайков или постинг (пользователи нанимаются для создания контента маркетингового характера). Все это только подчеркивает сложность отбора методик, которые позволят объективно оценить пользовательскую активность.

В 2021 г. в рамках программы стратегического академического лидерства «Приоритет 2030» по инициативе факультета социальных наук Национального Исследовательского Нижегородского государственного университета им. Н.И. Лобачевского была открыта лаборатория «Дизайн информации», ориентированная на создание комплексной программы исследования уровня креативности студентов на основе психологических и социологических методик, а также на основе интеллектуального анализа пользовательских данных в социальных сетях. Первая проблема, с которой столкнулся научный коллектив, стал выбор валидных методик оценки предмета изучения по цифровым следам.

В целях оценки релевантных методик исследования креативности средствами Big Data был проведен контент-анализ статей из рейтинговых зарубежных журналов, посвященных оценке качеств личности по цифровым следам. Выборка составила 50 статей с высокими показателями цитируемости из библиографической, реферативной базы Scopus, за период с 2001 по 2021 г. Для обработки полученных данных был применен метод контент-анализа с привлечением специализированного пакета Лекта. Базовый словарь составил более 12 086 слов.

Анализ зарубежных источников показывает, что есть три крупных направления исследований, где по цифровым следам определялся:

- социально-демографический портрет пользователя: пол, возраст и территория проживания, сексуальная ориентация, этническая принадлежность, религиозные взгляды;
- «Пятифакторная модель личности» («Большая пятерка» или «Big five»: доброжелательность, экстраверсия, открытость к новому, самоконтроль, нейротизм);
- рост уровня социальных проблем, например, заболеваемости гриппом, безработицы, изменение отношения пользователей к мигрантам.

Зарубежные исследователи используют разнообразные источники данных для сбора Big Data. Самым популярным является анализ постов. Можно выделить три варианта их изучения:

– лингвистический анализ постов: например, выявлено, что люди с более высокими показателями макиавеллизма склонны использовать больше бранных и оскорбительных слов [9, 10];

– анализ стиля написания поста (количество вопросительных предложений, особенности пунктуации, употребление местоимений, выбор частей речи, междометия и т.д.). Например, по этим характеристикам выявляется склонность человека к сотрудничеству и самоконтролю [11, 12];

– выбор темы сообщения: например, пользователи, упоминающие «Hello Kitty» склонны к высоким показателям открытости и низким показателям самоконтроля [13, 14].

Предсказательную силу имеет степень заполненности профиля страницы в социальных сетях [15] и частота обновления статуса [16]: например, самоконтроль отрицательно коррелирует с членством в группах, и положительно – с количеством загруженных фотографий. Люди, склонные к сотрудничеству, больше заполняют свой профиль, и используют дополнительный функционал страниц в социальных сетях [16].

Данные о лайках постов позволяют предсказывать этническое происхождение, религиозную принадлежность, сексуальную ориентацию, и некоторые черты характера. Коллективом авторов под руководством М. Косински разработаны матрицы «привлекательных» тем, соотнесённые с такими чертами, как экстраверсия, сотрудничество, самоконтроль, нейротизм, и открытость новому опыту [14].

Структура и плотность дружеских сетей – также материал для анализа. Например, несколькими исследованиями доказано, что экстраверсия и эгоцентризм коррелируют с количеством контактов в Facebook [17–19].

Сложно изучаемым, но информативным источником данных являются фотографии, размещаемые пользователями в социальной сети: содержание, теги, название, даты загрузки, комментарии (отзывы). Несмотря на дискуссии о том, какие фото чаще размещаются – того, что привлекает, или того, что есть в реальной жизни, выявлено, что фото своего дома в широком смысле, размещаются чаще, чем другие места [20]. Экстравертивные люди выкладывают фото, где улыбаются и изображены с другими людьми, интроверты, как правило, появляются в одиночестве, а невротики предпочитают изображения без людей или лица крупным планом [21].

Источником данных о пользователе может быть даже выбор музыкальных подписок: среди людей, предпочитающих классическую музыку, больше изобретателей, они также достаточно терпимы [22].

Несмотря на широкие перспективы использования цифровых следов в современных научных исследованиях, все авторы отмечают общие проблемы: сложность в доказательстве достоверности получаемых данных, большой объем нерелевантного контента, искаженный или неоднозначный смысл сообщений, влияние ситуативных факторов на поведение в сети, несформировавшуюся позицию пользователя, а, главное, отсутствие принятых в научном сообществе верифицированных методик анализа полученных массивов данных.

Таким образом, опыт зарубежных исследований, проведенных на основе Big Data, показывает возможности определения широкого спектра личностных качеств по цифровым следам. Однако следует учитывать, что некоторые черты личности адекватно отражаются в разных цифровых записях: одни – верифицируемы в переписке, другие – в графических материалах, третьи – в музыкальных предпочтениях и т.д., что существенно затрудняет работу исследователя. Материалов, посвященных изучению креативности, выявлено не было.

Кейсы иностранных ученых показывают, что интернет-дискурс – это пространство для выражения не только виртуальных, идеализированных, но и реальных личностных характеристик, поэтому прогнозирование индивидуальных черт и предпочтений пользователей – актуальный вызов междисциплинарным исследованиям.

Список источников

1. Волков В.В. Технологизация социологии // Мониторинг общественного мнения: экономические и социальные перемены. 2004. № 4 (72). С. 80–81.
2. Кузина О.Е., Шашкин А.В. Сравнительный анализ методов квартирного и он-лайн опросов при изучении рынка финансовых услуг // Мониторинг общественного мнения: экономические и социальные перемены. 2009. № 2 (90). С. 172–178.
3. Ильясов Ф.Н. Массовое сознание россиян в зеркале Яндекса // Мониторинг общественного мнения: экономические и социальные перемены. 2009. № 4 (92). С. 150–156.
4. Азаров А.А., Бродовская Е.В., Дмитриева О.В. [и др.]. Стратегии формирования установок протестного поведения в сети интернет: опыт применения киберметрического анализа (на примере Евромайдана, ноябрь 2013 г.) // Мониторинг общественного мнения: экономические и социальные перемены. 2014. № 2 (120). С. 63–78.
5. Колесниченко О.Ю., Смородин Г.Н., Ильин И.В. [и др.]. Многоцентровое исследование по аналитике BIG DATA Академического партнерства ЕМС в России Третья волна // Мониторинг общественного мнения: экономические и социальные перемены. 2015. № 5 (129). С. 21–41.
6. Дудина В.И., Юдина Д.И. Извлекая мнения из сети интернет: могут ли методы анализа текстов заменить опросы общественного мнения? // Мониторинг общественного мнения: экономические и социальные перемены. 2017. № 5 (141). С. 63–78.
7. Щекотин Е.В., Мягков М.Г., Гойко В.Л. [и др.]. Субъективная оценка (не)благополучия населения регионов РФ на основе данных социальных сетей // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 1 (155). С. 78–116.
8. Бразевич С.С., Сафронова Ж.С., Бразевич Д.С., Матюха Д.В. Исследование интересов молодежи через тематический репертуар общения в студенческой среде: кейс социальной сети «ВКонтакте» // Мониторинг общественного мнения: экономические и социальные перемены. 2019. № 1 (149). С. 193–211.
9. Sumner C., Byers A., Booschever R., Park G.J. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. Machine learning and applications (ICMLA) // Proceedings of the 11th International conference on Machine learning and applications. IEEE, 2012. P. 386–393.
10. Wald R., Khoshgoftaar T., Sumner C. Machine prediction of personality from Facebook profiles // Proceedings of the 13th International Conference on Information Reuse and Integration. IPI, 2012. P. 109–115.
11. Pervaz I., Ameer I., Sittar A, Nawab RMA. Identification of author personality traits using stylistic features // CLEF. 2015. P. 1–8.
12. Przybyła P, Teisseyre P. What do your look-alikes say about you? Exploiting strong and weak similarities for author profiling // CEUR Workshop Proceedings. 2015. P. 1–8.
13. Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior // Proceedings of the National Academy of Sciences. 2013. № 110 (15). P. 5802–5805.
14. Qiu L., Lin H., Ramsay J., Yang F. You are what you tweet: Personality expression and perception on Twitter // Journal of Research in Personality. 2012. № 46 (6). P. 710–718.
15. Amichai-Hamburger Y., Vinitzky G. Social network use and personality // Computers in Human Behavior. 2010. № 26 (6). P. 1289–1295.
16. Anderson C., John O.P., Keltner D., Kring A.M. Who attains social status? Effects of personality and physical attractiveness in social groups // Journal of Personality and Social Psychology. 2001. № 81 (1). P. 114–116.
17. Quercia D., Lambiotte R., Stillwell D., Kosinski M., Crowcroft J. The personality of popular Facebook users // Proceedings of the ACM Conference on Computer Supported Cooperative Work. CSCW, 2012. P. 955–964.
18. Markovikj D., Gievska S., Kosinski M., Stillwell D. Mining Facebook data for predictive personality modeling // Proceedings of the 7th international AAAI conference on weblogs and social media (ICWSM 2013). Boston, MA, USA, 2013. P. 23–26.
19. Quercia D., Lambiotte R., Stillwell D., Kosinski M., Crowcroft J. The personality of popular Facebook users // Proceedings of the ACM 2012 conference on computer supported cooperative work. ACM, 2012. P. 955–964.
20. Popescu A., Grefenstette G. Mining User Home Location and Gender from Flickr Tags // Proceedings of the 4th International AAAI Conference on Weblogs and Social Media. ICWSM, 2010. P. 307–310.
21. Celli F., Bruni E., Lepri B. Automatic personality and interaction style recognition from Facebook profile pictures // Proceedings of the 22nd ACM international conference on multimedia. ACM, 2014. P. 1101–1104.
22. Rentfrow P.J., Gosling S.D. The do re mi's of everyday life: The structure and personality correlates of music preferences // Journal of Personality and Social Psychology. 2003. № 84 (6). P. 1236–1256.

POSSIBILITIES OF CREATIVITY ASSESSMENT BY DIGITAL FOOTPRINTS

I.A. Isakova

*Lobachevsky State University of Nizhny Novgorod – National Research University,
Nizhny Novgorod, Russian Federation*

Abstract. The article analyzes foreign experience in identifying personality traits through digital footprints in social networks. The study is conducted in the form of a content analysis of articles from top foreign journals on the assessment of personality traits using digital footprints. The sample consisted of 50 articles with high citation rates from the bibliographic reference database Scopus for the period from 2001 to 2021. Digital footprints are understood as a totality of data that a person consciously and unknowingly leaves behind when using the Internet. As a result, positive and negative aspects of using digital footprints for a social-psychological profile of users are described, as well as methods of assessing personality traits according to the digital footprints. The article also presents the experience of creativity research using Big Data in the Information Design laboratory of Lobachevsky State University of Nizhny Novgorod – National Research University.

Keywords: digitalization, creativity, digital footprint, content analysis, Internet, personal qualities, user

ПРАГМАТИКА ДЕСИГНАТОВ БЛАГОТВОРИТЕЛЬНОСТИ: ОПЫТ СЕТЕВОГО АНАЛИЗА АНГЛОЯЗЫЧНЫХ НАУЧНЫХ ТЕКСТОВ

В.В. Кашпур, А.А. Барышев, А.В. Губанов, Е.С. Мищенко

Национальный исследовательский Томский государственный университет, Томск, Россия

В данной работе представлен опыт анализа прагматики десигнатов социального концепта «благотворительность» в современных англоязычных научных текстах (2020–2022 гг.). Методом анализа дефиниций были определены единицы, вербализующие десигнаты благотворительности – *charity* и *philanthropy*. Методом библиометрического анализа были выявлены и визуализированы контекстные связи данных единиц. Исходная база данных для построения библиометрической сети включает около 6000 статей (названия и аннотации), выгруженных из базы данных Web of Science и Scopus. Для каждой из анализируемых единиц была построена сеть взаимовстречаемости слов. В результате анализа этих сетей была верифицирована исследовательская гипотеза о том, что употребление лексических единиц *charity* и *philanthropy*, традиционно воспринимаемых как взаимозаменяемые синонимы, формирует специфические семантические поля десигнатов анализируемого социального концепта.

Ключевые слова: большие данные, научный текст, благотворительность, библиометрический анализ

Источник финансирования: исследование выполнено при поддержке Программы развития ТГУ («Приоритет-2030»).

Благотворительность – социальный концепт, который формируется и изменяется в процессе говорения и думания. В основе благотворительной деятельности лежат такие категории как альтруизм, гуманизм, сострадание и бескорыстие. Определение благотворительности является процессом постоянным (за счёт наращивания значений) и социальным (определяется участниками разговоров о благотворительности). Именно во взаимодействии акторов (учёных, представителей релевантных организаций, политических структур, масс-медиа и т.д.) происходит конструирование благотворительности как социального феномена. Поэтому релевантным подходом к исследованию благотворительности является социально-конструкционистский, согласно которому в формировании социальной реальности (и ее феноменов) активную роль играют речевые практики (дискурсы) [1]. Так как разные говорящие в своих речевых практиках актуализируют разные (различающиеся) значения благотворительности, социальный концепт «благотворительность» не имеет однозначного определения, и изучение способов номинации концепта и десигнатов, которые эти номинации вербализуют, представляется актуальным.

Исследование прагматики десигнатов благотворительности предполагает их изучение в контексте – как значения языковых единиц, эксплицирующих концепт «благотворительность», возникающих в процессе их употребления [2, 3]. Соответственно, анализ контекстных связей единиц, вербализующих десигнаты концепта, позволяет выявить прагматику значений языковых единиц – формируемые говорящими семантические поля значений концепта.

Цель исследования – сочетая методы лингвистического и библиометрического анализа текста, определить семантические поля, формируемые номинациями десигнатов социального концепта «благотворительность» в англоязычных научных текстах.

Представляемое исследование является междисциплинарным. Для выявления единиц для анализа был использован лингвистический лексикографический подход (анализ дефиниций); для выявления и визуализации связей между словами, эксплицирующими десигнаты благотворительности, был применён библиометрический сетевой анализ; для верификации выводов автоматической обработки текстов был применён «ручной» анализ выявленных связей (Google Ngram Viewer, Google Scholar).

На основе анализа предлагаемых русско-английскими словарями [4, 5] переводов русскоязычной единицы «благотворительность» было выявлено 18 вариантов ее перевода на английский язык. В результате анализа значений этих вариантов по моноязычным английским словарям [6, 7] были определены две ключевые единицы, выражающие десигнаты благотворительности – *charity* и *philanthropy*.

Анализ словарных дефиниций позволил выявить семантические центры в лексических единицах. Так, «*charity*» – это организация или комплекс организаций, оказывающие прежде всего денежную помощь, осуществляющие сборы средств, принимающие пожертвования, а также другие виды помощи **нуждающимся**. «*Philanthropy*» – (абстрактное) желание способствовать благополучию людей (сема нуждающихся не эксплицирована в словарном определении), северно-американское употребление слова соответствует значению «*charity*». Выявленное наложение значений объясняет традиционное восприятие единиц *charity* и *philanthropy* как взаимозаменяемых синонимов.

На следующем этапе работы над анализом прагматики единиц, выражающих десигнаты социального концепта «благотворительность», применяя инструмент библиометрического сетевого анализа, была произведена автоматическая обработка базы текстов с целью создания сети совместной встречаемости слов, позволяющей визуализировать их взаимосвязанность.

Библиометрический анализ проводился на основе базы данных, сформированной авторами данной статьи. Поиск источников для исследования производился в базе данных Web of Science по ключевым словам (*charity, non-profit, fundraising* и т.д.). Результаты поиска были отсортированы по соответствию и выгружены. В результате было получено 5 343 публикации. Информация, которая доступна при выгрузке: ФИО автора(-ов), название публикации, данные о журнале, ключевые слова, аннотация. С целью дополнения данных также были выгружены статьи из базы данных Scopus с ключевым словом *philanthropy* с 2020 по 2022 г. Поиск производился по названиям и ключевым словам, а также отфильтрован по научным областям: искусство и гуманитарные науки (ARTS), бизнес, управление и бухгалтерский учет (BUSI), науки о принятии решений (DECI), экономика, эконометрика и финансы (ECON), психология (PSYC) и социальные науки (SOCI). В результате было выявлено 520 статей. Для построения библиометрической сети были использованы названия и абстракты выгруженных статей.

Так как данная работа направлена на изучение контекста, в котором употребляются слова *charity* и *philanthropy*, для каждого из этих слов была построена его эго-сеть (сеть из тех слов, которые употребляются вместе с ним). Построение сети производилось в несколько этапов. На первом этапе все тексты были токенизированы и лемматизированы; некоторые важные для исследования термины с разными традициями написания были приведены к единой форме (как, например, *organisation* или *organization, well-being, wellbeing* или *well being*). Далее из сети были убраны все тексты кроме тех, которые содержат ключевое слово (*charity* или *philanthropy*). На следующем этапе из всех текстов были удалены стоп-слова, как общеупотребительные (*any, other, and*), так и специфичные для научных статей (*research, study, result*). Из оставшихся слов были выбраны единицы, которые встречаются хотя бы в 6 документах, и на полученном материале была построена сеть. Отметим, что само ключевое слово (*charity* или *philanthropy*) из сети удаляется, так как оно связано со всеми остальными единицами и в рамках сети не несёт никакой смысловой нагрузки. В построенной сети могут обнаружиться новые стоп-слова, поэтому список стоп-слов пополняется, в результате чего сеть перестраивается. После нескольких итераций сеть приходит к виду, пригодному для анализа семантических полей в интересующей нас предметной области.

На рис. 1 представлена сеть взаимовстречаемости слов единицы «*charity*».

В данной сети отчётливо выделяются два кластера с основными узлами «*charitable*» и «*social*». В кластере с узлом «*charitable*» визуализированные связи слов позволяют сделать ряд наблюдений о специфике эксплицируемых значений «*charity*»:

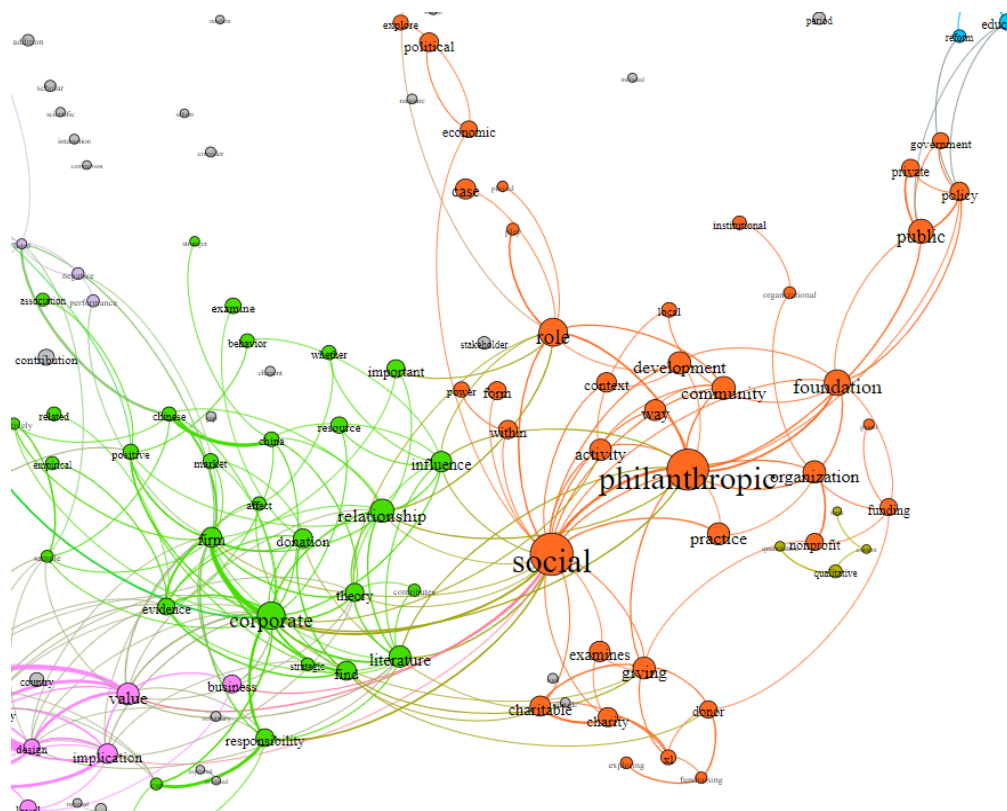


Рис. 2. «Philanthropy»: сеть взаимовстречаемости слов

В рамках первого кластера актуализируются следующие значения *philanthropy*:

1) связи слова *philanthropic* со словами *practice*, *activity*, а также *social* с *practice*, акцентируют социальный и деятельностный аспекты благотворительности, как и в сети «charity»;

2) связи слова *philanthropic* со словами *role*, *social*, *influence*, а также *social* с *economic*, *political*, *role* эксплицируют роль и влияние благотворительности в социальном мире; данное семантическое поле отсутствует в сети «charity»;

3) связи слова *philanthropic* со словами *development*, *local*, *community* характеризуют роль благотворительности как способа развития некоего сообщества; значимым представляется выбор слова для обозначения объекта благотворительности – *community* как сообщества людей, объединённых на основе какого-либо признака (не индивида или людей вообще, как в сети «charity»), а сознательно выделенной на основе определённых критериев группы);

4) связи слова *philanthropic* со словами *foundation* (в свою очередь связанные с единицами *public*, *private*, *government*, *policy*) и *organization* (*non-profit*, *funding*, *grant*) выражают значение институциональной оформленности благотворительных действий.

Отметим наличие в сети «philanthropy» небольшого кластера слов, также принадлежащих сети *charity*: *charitable*, *charity*, *donor*, что свидетельствует о пересечении не только значений двух единиц (*philanthropy* и *charity*), но и способов их номинации (обратной тенденции выявлено не было: слово *philanthropy* и его производные отсутствует в сети *charity*). Эти слова (*charitable*, *charity*, *donor*) и их связи со словами *social* и *philanthropic giving* отражают значения, связанные с содержательным и организационным аспектами осуществления благотворительности.

Связь первого кластера (с узлами *philanthropic* и *social*) со вторым (*corporate*) осуществляется через слова *social*, *corporate*, *responsibility*, *influence*; эксплицируются значения благотворительности как социального института (не только фонды, но и корпорации являются участниками благотворительности), существующего и играющего свою роль по опреде-

лённым правилам (корпоративная социальная ответственность традиционно фиксируется в уставных документах компаний).

Об институционализации благотворительности в сети «philanthropy» свидетельствуют связи, выявленные во втором кластере.

Связь единицы *corporate* со словом *firm* означает организацию с программой осуществления благотворительности, с *responsibility, ethics* – корпоративную социальную ответственность (выражающуюся в том числе и в благотворительной деятельности), закреплённую в миссии организации. В кластере присутствует узел *CSR (corporate social responsibility)*, наличие акронима свидетельствует о традиционности и закреплённости данного значения.

Таким образом, в сети «philanthropy» представлены семантические поля социальной деятельностью природы благотворительности, ее общественной значимости, объекта благотворительности (выделяемые по неким критериям группы людей), институциональности и внутреннего (корпоративного) регулирования благотворительной деятельности. Единица *philanthropy* используется при описании системной долгосрочной стратегической программы оказания помощи институтами не случайным объектам, а критериально определенным сообществам людей.

В таблице отражены основные особенности употреблений *charity* и *philanthropy*.

Charity и philanthropy: специфика прагматики десигнатов

Семантическое поле	CHARITY	PHILANTHROPY
Характер деятельности	отдельный акт + регулярная работа <i>charitable donation, giving; charitable practice/work</i>	регулярная работа + программа <i>philanthropic practice/work; philanthropic funding</i>
Тип организации	благотворительная организация (фонд) <i>charitable non-profit organization</i>	благотворительная организация (фонд) + корпорация <i>philanthropic organization/foundation; corporate donation</i>
Объект	индивид <i>individual</i>	сообщество <i>community</i>

Номинация *philanthropy* включает в себя основные значения, эксплицируемые номинацией *charity*, что приводит к избыточности употребления последней, с одной стороны, но одновременно заставляет её активизировать креативный потенциал для компенсации содержательных лакун, что приводит к созданию новых семантических полей: в нашей сети *charity* это иллюстрируется наличием изолированного кластера *digital charity*.

Таким образом, анализ взаимодействий слов, вербализующих и контекстно связанных со значениями благотворительности, продемонстрировал возможность выявления семантических полей, образуемых этими словами, что способствует более глубокому пониманию системы значений концепта и реализуемых этими значениями прагматических установок. Дальнейшая работа по увеличению базы данных, ее чистке позволит верифицировать, расширить, уточнить полученные на данном этапе работы выводы.

Список источников

1. Бергер П., Лукман Т. Социальное конструирование реальности: Трактат по социологии знания / пер. с англ. Е. Руткевич. М.: Academia-Центр; Медиум, 1995. 323 с.
2. Leech G.N. Principles of Pragmatics. Longman, 1983. 250 p.
3. Levinson S.C. Pragmatics. Cambridge: Cambridge University Press, 1983. 438 p.
4. Мюллер В.К. Большой англо-русский и русско-английский словарь. 450 000 слов и словосочетаний. Новая редакция. Москва: Славянский Дом Книги, 2015. 960 с.
5. Словарь «Mltitran». URL: <https://www.mltitran.com> (дата обращения: 08.05.2022).
6. Google's English Dictionary // Oxford Languages. URL: <https://languages.oup.com/google-dictionary-en/> (дата обращения: 08.05.2022).
7. Collins English Dictionary. URL: <https://www.collinsdictionary.com/> (дата обращения: 08.05.2022).

**PRAGMATICS OF CHARITY/PHILANTHROPY DESIGNATA: AN EXPERIENCE
OF NETWORK ANALYSIS OF ENGLISH-LANGUAGE ACADEMIC TEXTS**

V.V. Kashpur, A.A. Baryshev, A.V. Gubanov, E.S. Mishchenko

National Research Tomsk State University, Tomsk, Russian Federation

Abstract. This paper presents the experience of analyzing the pragmatics of the designata of the social concept *blagotvoritel'nost'* (charity, philanthropy, etc.) in contemporary English-language academic texts (yy. 2020–2022). Using definition analysis, words were identified that verbalize the designata of *blagotvoritel'nost'* – “charity” and “philanthropy”. Using bibliometric analysis, these words’ contextual relationships were identified and visualized. The initial database for building a bibliometric network includes about 6000 articles (titles and abstracts) retrieved from the Web of Science and Scopus database. For each of the analyzed units, a network of word cooccurrences was built. As a result of the analysis of the networks, the research hypothesis was verified that the use of the lexical units “charity” and “philanthropy”, traditionally perceived as interchangeable synonyms, forms specific semantic fields of the designata of *blagotvoritel'nost'*.

Keywords: big data, academic text, charity, philanthropy, bibliometric analysis

О ЦИФРОВОМ АССИСТЕНТЕ ДЛЯ МОНИТОРИНГА ЖИЗНЕСТОЙКОСТИ ЧЕЛОВЕКА В УСЛОВИЯХ ПОВСЕДНЕВНОЙ ЖИЗНИ

Д.Ж. Корзун

Петрозаводский государственный университет, Петрозаводск, Россия

Представлены результаты исследования бионических методов сенсорики и окружающего искусственного интеллекта для обеспечения жизнестойкости человека в условиях повседневной жизни на северных территориях. Исследуемые методы поддерживают разработку интернет-сред мониторинга жизнестойкости человека преимущественно в сфере ментального здоровья. Представлена концептуальная модель цифрового ассистента человека, позволяющая выполнять регулярный сбор данных о процессах, происходящих при выполнении человеком автономной, когнитивной и двигательной функций. Обсуждается понятие «умный бионический костюм», в виде которого может быть реализован цифровой ассистент. Поставлены задачи по построению моделей и алгоритмов, обеспечивающих многослойную архитектуру обработки данных в интернет-системе мониторинга жизнестойкости человека в режиме реального времени.

Ключевые слова: жизнестойкость человека, окружающий искусственный интеллект, бионические методы сенсорики, интернет-системы, виртуальная и дополненная реальность

Источник финансирования: исследование выполняется при поддержке гранта Российского научного фонда № 22-11-20040 (<https://rscf.ru/project/22-11-20040/>), проводимого совместно с Республикой Карелия с финансированием из Фонда венчурных инвестиций Республики Карелия (ФВИ РК).

Цифровой ассистент человека (ЦАЧ)

Предлагается построить вокруг человека интернет-среду мониторинга (ИСМ) [1], которая способна сопровождать деятельность человека в условиях его повседневной жизни. Сопровождение подразумевает регулярный сбор данных о процессах, происходящих при выполнении человеком автономной (работы организма), когнитивной (восприятие и познание) и двигательных (физическая активность) функций. На основе обработки этих данных выполняется построение информационных сервисов, усиливающих органы чувств человека. В общем случае такие сервисы позволяют человеку «заглянуть» в происходящие с ним процессы.

В качестве референтного варианта рассматривается ЦАЧ для мониторинга жизнестойкости человека. Под «жизнестойкостью» понимается позитивная установка личности (и ее составляющие), позволяющая человеку успешно преодолевать жизненные трудности, в чем бы они ни заключались (болезнь, стресс, экстремальные ситуации, конкретные жизненные проблемы и утраты, однообразность и рутинность повседневности и пр.) [2]. ЦАЧ должен помочь человеку осознано и целенаправленно воспринимать свою жизненную ситуацию и находить ее решение. Исследуются следующие проявления жизнестойкости человека.

1. Сопrotивляемость ментального здоровья стрессам и другим негативным внешним факторам. Помощь человеку в своевременном понимании текущей ситуации, причин возникновения, рекомендации к преодолению.

2. Адаптация к условиям работы и проживания. Помощь человеку, перешедшему на новую работу или приехавшему на новое место жительства.

3. Сопrotивляемость усталости. Отслеживание состояния человека с целью распознавания снижения эффективности работы, предупреждение нарушений, обеспечение охраны труда и здоровья.

4. Мотивация к сохранению здоровья. Помощь человеку в следовании правилам здорового образа жизни, в использовании спортивного оборудования, ведение дневника с распознаванием успехов и неудач.

В условиях северных территорий жизнь человека подвержена воздействию большого объема негативных факторов [2]. В то же время люди, проживающие на таких территориях, накопили своеобразный опыт, обеспечивающий их повышенную жизнестойкость (например, холодоустойчивое поведение). Этот опыт может быть интегрирован в рассматриваемый референтный ЦАЧ.

Возможности цифрового ассистента

Выделим следующие модельные ситуации для использования ЦАЧ в условиях повседневной жизни на северных территориях.

- Ситуация «ЮС-переезд». Люди, приезжающие на север с южных территорий. Пример – студенты, приезжающие в университет на учебу.

- Ситуация «Малолюдность». Выполнение работ в одиночку на малонаселенной территории, т.е. низкое число очных контактов с другими людьми. Другой пример – дистанционная работа (оператор, диспетчер, аналитик др.).

- Ситуация «Ограниченность». Человек с ограниченными возможностями, в том числе с ограничениями когнитивной и двигательной функций вследствие перенесенных болезней и других нарушений здоровья. Другой пример – мониторинг пациентов и пожилых людей.

Такие ситуации возникают, когда, например, человек:

- работает за компьютером (рабочее место, дистанционная работа, работа дома);
- передвигается по территории (в помещении, на улице, на производстве);
- занимается спортом (на тренажере).

При этом обеспечение жизнестойкости не делегируется ЦАЧ, а от человека требуется самопознание. За счет технологий сенсорики можно «заглянуть в себя», использовать дополнительные к органам чувств источники данных. В свою очередь, методы искусственного интеллекта (на основе анализа таких данных) могут помочь человеку ответить на вопросы: «Что происходит с моим организмом и здоровьем?», «Как я двигаюсь и что происходит вокруг?», «Как я воспринимаю происходящее?».

Понятие «умный бионический костюм»

ИСМ (реализующая ЦАЧ) состоит из множества взаимодействующих вычислительных устройств (разнообразные компьютеры, в том числе датчики и человеко-машинные интерфейсы). Такие устройства, во-первых, обеспечивают сенсорику для следующих видов данных (пересекающиеся): а) самочувствие, б) холодоустойчивость и другие факторы внешней среды, в) общие параметры крови, г) вовлеченность в выполняемую деятельность, д) тревожность, е) скорость реакции, ж) параметры движения. Во-вторых, необходима визуализация и другие формы представления результатов анализа в (псевдо)реальном времени для восприятия человеком, преимущественно, в стиле «объект мониторинга рассказывает о себе».

Требуется построение ИСМ на основе следующих типов носимых, нательных и встроенных вычислительных устройств, в первую очередь, мобильных.

- Бытовая сенсорика здоровья: браслеты, нательные датчики, смартфоны, умный текстиль.

- Цифровое окружение человека: видеокамеры и другие сенсорные системы.

Следует отметить, что в лабораторных медицинских условиях используется сертифицированное высокоточное измерительное оборудование. Оно неприменимо для рассматриваемого варианта ЦАЧ для мониторинга жизнестойкости человека, так как не поддерживает требования мобильности и обработки в реальном времени [3]. Многие датчики пока являются лишь перспективными разработками (см., например, [4, 5]) для задач анализа крови. Обзор разнообразных IoT-датчиков для сбора данных о состоянии здоровья человека представлен в [6, 7].

С учетом вышеприведенных модельных ситуаций «ЮС-переезд», «Малолюдность» и «Ограниченность», необходимо обеспечивать свойства мобильности и автономности, то есть система сопровождает человека и не гарантируется постоянный доступ к глобальному интернету. Эти свойства достигаются за счет бионических принципов для построения ИКТ-систем. Во-первых, сенсорика использует бионические датчики для отслеживания движения [8], которые могут быть встроены в одежду (умный текстиль), приходя к понятию «бионического костюма» [9].

Во-вторых, получение результатов анализа и доставка их для восприятия человеком также использует бионический принцип. В частности, возможно использование 2D и 3D моделей для визуализации процессов, происходящих с человеком, в том числе на основе технологий дополненной (AR) и виртуальной (VR) реальности. Например, используя шлем виртуальной реальности или умные очки, человек может наблюдать за своим движением по помещению.

В целом такую ИСМ для реализации ЦАЧ следует рассматривать как AmI-систему с многослойной архитектурой [1].

1. Слой сенсорики данных. Проблема – разнообразие IoT-датчиков.

2. Слой предобработки и накопления данных. Проблема – организация вычислений на имеющихся локально IoT-устройствах (периферийные вычисления – от англ. edge computing).

3. Слой семантического анализа данных (связывание). Проблема – использование графовых баз данных с механизмами быстрых поисковых запросов, ранжирования информации и построения рекомендаций.

4. Слой построения информационных сервисов.

5. Слой доставки информационных сервисов. Проблема – реализация быстрой обратной связи, когда человек наблюдает происходящие процессы в режиме, близком к реальному времени.

Реализация функциональности этих слоев очевидно приводит к проблеме больших данных, так как поступающие от датчиков данные имеют потоковый характер, могут обладать высокой интенсивностью, а мониторинг выполняется непрерывно.

Таким образом, для выполнения указанных свойств мобильности, автономности и обработки множества потоков данных в режиме реального времени на указанных слоях требуется построение специализированных математических моделей и распределенных алгоритмов. Для разработки требуемого ЦАЧ выделим следующие ключевые задачи по построению моделей и алгоритмов.

- Модели и алгоритмы для предобработки «быстрых» потоковых сенсорных данных с управлением баланса «точность/трудоемкость».

- Адаптивные стратегии доступа к приоритетной сенсорной информации из множества источников для организации быстрой обратной связи.

- Модели и алгоритмы обработки многопараметрических данных для бионического костюма с применением технологий виртуальной и дополненной реальности для быстрого восприятия информации.

Заключение

В работе предложено построить вокруг человека ИСМ на базе разнообразных IoT-устройств. Реализуется ЦАЧ для мониторинга жизнестойкости человека, когда информационные сервисы строятся на основе обработки поступающих сенсорных данных. Выделены ключевые задачи для дальнейшего научного исследования, связанные с:

а) предобработкой «быстрых» потоковых сенсорных данных с управлением баланса «точность/трудоемкость»;

б) адаптивным доступом к приоритетной сенсорной информации из множества источников для организации быстрой обратной связи;

в) обработкой многопараметрических данных с применением технологий виртуальной и дополненной реальности.

Решение этих задач позволит получить новые бионические методы сенсорики и окружающего искусственного интеллекта для обеспечения жизнестойкости человека.

Список источников

1. Korzun D., Balandina E., Kashevnik A., Balandin S., Viola F. Ambient Intelligence Services in IoT Environments: Emerging Research and Opportunities. USA: IGI Global, 2019. 166 p.
2. Воробьева Н.А., Трошина Т.И. Человек и Север: к вопросу о социокультурных и медико-биологических источниках жизнестойкости // Экология человека. 2019. № 7. С. 42–48. doi: 10.33396/1728-0869-2019-7-42-48
3. Meigal A.Yu. et al. Ambient Assisted Living At-Home Laboratory for Motor Status Diagnostics in Parkinson's Disease Patients and Aged People // Tools and Technologies for the Development of Cyber-Physical Systems. 2020. doi: 10.4018/978-1-7998-1974-5.ch007
4. Vizbaras A. et al. GaSb Swept-Wavelength Lasers for Biomedical Sensing Applications // IEEE Journal of Selected Topics in Quantum Electronics. 2019. Vol. 25, № 6. P. 1–12. Art no. 1501812. doi: 10.1109/JSTQE.2019.2915967.
5. Le T. et al. Continuous Non-Invasive Blood Pressure Monitoring: A Methodological Review on Measurement Techniques // IEEE Access. 2020. Vol. 8. P. 212478–212498. doi: 10.1109/ACCESS.2020.3040257
6. Islam S.M.R., Kwak D., Kabir M.H., Hossain M., Kwak K.-S. The Internet of Things for Health Care: A Comprehensive Survey // IEEE Access. 2015. Vol. 3. P. 678–708. doi: 10.1109/ACCESS.2015.2437951
7. Aledhari M., Razzak R., Qolomany B., Al-Fuqaha A., Saeed F. Biomedical IoT: Enabling Technologies, Architectural Elements, Challenges, and Future Directions // IEEE Access. 2022. Vol. 10. P. 31306–31339. doi: 10.1109/ACCESS.2022.3159235
8. Рентюк В. Рука об руку с человеком: бионический манипулятор от FESTO // Control Engineering Россия. 2019. № 2 (80). С. 50–55.
9. Смирнов К.А., Ермаков В.А. Умный текстиль в разрезе тактильного интернета: к разработке системы мониторинга жизнестойкости человека в условиях северных территорий // Путь в науку: прикладная математика, информатика и информационные технологии : тезисы докладов Всероссийской молодёжной научно-практич. конф. Ярославль : Ярославский государственный университет им. П.Г. Демидова, 2022. С. 15–18.

ON A DIGITAL ASSISTANT FOR MONITORING HUMAN RESILIENCE IN EVERYDAY LIFE

D.G. Korzun

Petrozavodsk State University, Petrozavodsk, Russian Federation

Abstract. The work presents an ongoing research of bionic methods of sensorics and surrounding artificial intelligence to ensure human resilience in the conditions of everyday life in the northern territories. The methods under study support the development of online environments for monitoring human resilience, mainly in respect to mental health. A concept model of a human digital assistant is presented, which supports regular data collection for ongoing processes when a person performs autonomous, cognitive and motion functions. The concept of a smart bionic suit is discussed towards the form of digital assistant implementation. The problem is set for constructing models and algorithms that provide a multi-layer architecture of data processing in an Internet system for monitoring human resilience in the real-time mode.

Keywords: human resilience, ambient intelligence, bionic methods in sensorics, internet systems, virtual and augmented reality

БОЛЬШИЕ ДАННЫЕ В КОМПЬЮТЕРНОМ СПОРТЕ

Е.А. Косьмина

Национальный государственный Университет физической культуры, спорта и здоровья имени П.Ф. Лесгафта, Санкт-Петербург, Россия

В статье рассматриваются способы применения больших данных в спортивной отрасли. Приведены результаты исследований, являющиеся примерами успешной интеграции анализа больших данных в сферу спорта, с возможностью дальнейшего использования полученных результатов на практике. Компьютерный спорт как самый «цифровой вид спорта», результат спортивной деятельности которого можно измерить в единицах и нулях, является наиболее благоприятной средой для сбора и анализа больших данных, по сравнению с классическими видами спорта. Приведены конкретные примеры применения методов анализа «больших данных» в компьютерном спорте. Рассмотрена система определения читеров в «шутерах от первого лица», на основе «Байесовской сети». Обозначены возможные направления для исследований больших данных в компьютерном спорте. На основе анализа литературных источников предложен алгоритм анализа больших данных в компьютерном спорте.

Ключевые слова: большие данные, компьютерный спорт, киберспорт, интеллектуальный анализ данных, большие данные, машинное обучение

В настоящее время большое количество научных публикаций посвящено результатам и методам анализа больших данных. Предложенные методы все чаще используются в политике, маркетинге, науке. В сфере спорта анализ больших данных только начинает использоваться. В настоящее время основными направлениями использования технологии «big data» в спорте являются борьба с допингом, «договорными» матчами, анализ сердечной и мышечной деятельности спортсменов, компьютерные игры, оптимизация кадровой политики спортивных клубов [1].

К наиболее популярным методам анализа больших данных в настоящее время можно отнести: «data mining», машинное обучение, искусственные нейронные сети, «краудсорсинг», смешение и интеграцию данных, пространственный анализ, распознавание образов, прогнозную аналитику, имитационное моделирование, визуализацию аналитических данных, статистический анализ.

Отличным примером практического применения «Data Mining» в спортивной сфере является исследование Моисеева А.С. (2021), проведенное в Брянской области. Автор произвел анализ федерального государственного статистического наблюдения по форме 5-ФК «Сведения по организациям, осуществляющим спортивную подготовку» и выявил наиболее характерные взаимосвязи между квалификацией, наличием званий у тренерского состава и результатами выступления на всероссийских и международных соревнованиях. Предложенный метод позволил дать характеристику всем видам спорта, подвергшимся анализу, и оценить перспективы развития каждого вида спорта, прогнозировать «фундамент успеха» [2].

Интересным представляется разработка веб-сервиса поддержки принятия решений в селекционной работе волейбольного клуба Е.А. Лазебной (2019). Автором разработан сайт, на котором волейболист из любой точки мира и любой квалификации может оставить свою заявку о поиске команды. Тренеры имеют возможность просматривать и отбирать, наиболее перспективных на их взгляд игроков. С помощью приложения «Deductor» был проведен анализ более 10 000 заявок и выявлены наиболее значимые показатели для волейбольных тренеров при отборе перспективных спортсменов [3].

Современные спортивные технологии не стоят на месте и благодаря этому открываются все новые возможности для получения больших данных и их дальнейшего анализа. Так в настоящее время в NBA активно применяют новую систему камер «SportVU» для сбора статистической и тактической информации. Так, Н.А. Белоножкиной (2022) был проведен ана-

лиз информации, полученной с камер при проведении FIBA 3x3 World Tour 2019 г. В результате автором была составлена карта игровой площадки, разделенная на 14 секторов, с выявлением процента успешных бросков из каждого сектора [4].

Приведенные выше исследования, на наш взгляд, являются примерами успешной интеграции анализа больших данных в сферу спорта, с возможностью дальнейшего использования полученных результатов на практике.

Компьютерный спорт как самый «цифровой вид спорта», результат спортивной деятельности которого можно измерить в единицах и нулях, является наиболее благоприятной средой для сбора и анализа больших данных, по сравнению с классическими видами спорта. Любая киберспортивная дисциплина – это прежде всего компьютерная программа, позволяющая осуществлять сбор и хранение информации о спортсмене: параметры устройства (модель, размеры изображения, модель компьютерной мыши и клавиатуры и т.д.), язык, геолокация, информация о внутриигровых действиях (длительность, частота игры, выбор персонажей и т.д.). Особый интерес представляют поведенческие особенности игрока, обработка текстовых и голосовых чатов. По статистике пользователи онлайн игр генерируют более 50 терабайт данных в сутки [5]. Наибольшую сложность и ценность представляют систематизированные и проанализированные данные. Один из способов практического применения больших данных приведен в работе S. Yeung (2006). Автор разработал систему определения читеров в шутерах от первого лица, на основе «Байесовской сети». Теория Байеса – это главный инструмент машинного обучения. Байесовские сети (Bayesian network) – это один из самых удобных способов представлять большие и сложные распределения вероятностей, простыми словами это вероятностная модель, представляющая собой множество переменных и вероятность их зависимостей. В исследовании S. Yeung (2006) вероятностная модель выглядит так, игрок совершил точный выстрел, какова вероятность что он использовал чит-программу «aimbot». Сеть учитывала следующие переменные: двигался игрок во время стрельбы или стоял неподвижно, двигалась цель игрока или не двигалась, изменял игрок направление прицеливания или удерживал в одной точке, оценивал расстояние между игроком и целью. Все эти факторы влияют на точность стрельбы, чем ближе расстояние до цели, тем проще попасть, в неподвижную цель проще попасть чем в движущуюся, если игрок стоит на месте, то вероятность того, что он попадет в цель будет выше, чем если он будет двигаться. Таким образом наличие этих факторов позволяет определить вероятность того, что игрок использовал чит-программу. Однако авторы не остановились на этом алгоритме, а усложнили его, так как современные чит-программы умеют эмулировать поведение живого человека. Поэтому они анализировали три варианта бота. Первый вариант. Когда «aimbot» включен, он находит ближайшую цель и точно прицеливается в нее и будет продолжать целиться в текущую цель, даже при появлении более близкой цели, пока расстояние между игроком и текущей целью не превысит определенный порог. Если есть только одна видимая цель, «aimbot» будет продолжать целиться в эту цель до тех пор, пока цель не станет невидимой для игрока. Второй вариант. «Aimbot», автоматически включается и выключается в течение случайного интервала времени (от 0,5 до 2 с). Игрок временно берет на себя управление во время периодов отключения. Снижение точности прицеливания, снижает шанс обнаружения читера. Третий вариант. Выполняет промахи для некоторых случайных временных интервалов (от 0,5 до 2 с). Программа создает видимость, что промахивается игрок, демонстрируя плавное колебание прицела вокруг своей цели. На рис. 1 представлены графические результаты определения вероятности обмана, для честного игрока и трех вариантов бота [6].

Следующим примером результатов анализа «больших данных» в компьютерном спорте являются сайты, предоставляющие статистику по киберспортивным командам и игрокам. На подобных сайтах приводятся статистические данные по оценке: «текущей формы команды», т.е. рассчитывается по разнице побед и поражений за определенное количество игр, как правило 10–30, процент побед команды за сезон, могут быть приведены собственные коэффициенты, например на сайте sq.gu приводится SQ.Score – оценивающий текущий уровень игры

команды, который изменяется автоматически после каждого матча. В процессе каждого крупного киберспортивного турнира производится сбор, анализ и публикация полной статистики, приведем пример «The International» 2021, крупнейший турнир по Dota 2. В ходе турнира был произведен анализ по 114 игровым персонажам, рассчитывались: процент выбора персонажа за каждую игровую сторону (radiant, dire), процент побед и поражений, винрейт каждого персонажа, процент банов каждого персонажа, процент побед против определенных типов персонажей и наиболее эффективные сочетания с другими персонажами.

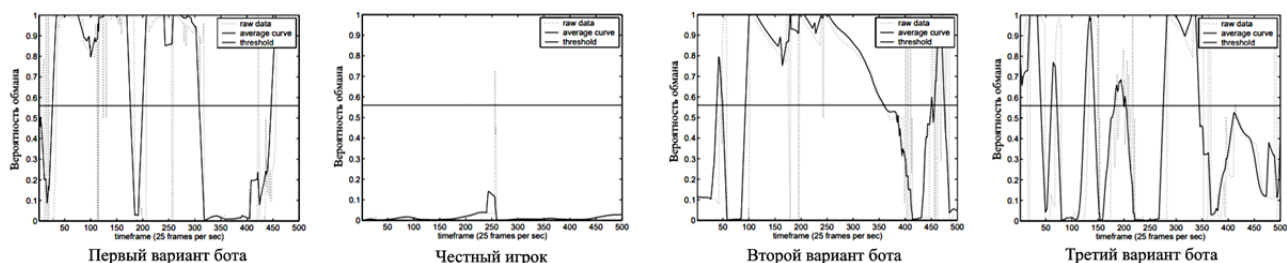


Рис. 1. Графическая оценка вероятности использования игроком чит-программ

В классических видах спорта, сбор, анализ и использование всех видов данных о деятельности спортсмена уже давно лежит в основе изменения понимания каждого вида спорта, тренеры, которые анализируют выступления соперников, профессиональных игроков выходят на новый уровень и меняют собственное понимание вида спорта.

В спортивной подготовке занимающихся компьютерным спортом анализ больших данных целесообразно применять в анализе поведения спортсмена или команды, для разработки алгоритмов технико-тактических действий. Представляется возможным проведение анализа расположения и перемещения игроков по игровой карте. Оценка эффективности спортсменов с целью последующего отбора. В компьютерном спорте, как и в классических видах спорта почти в каждой игровой дисциплине представлены системы аналитики (внутренние и внешние сервисы), анализ данных подобных сервисов на протяжении сезона может позволить провести анализ существующего состава игроков, выявив «слабые места» внести корректировки.

На основе анализа литературных источников нами предложен алгоритм анализа больших данных в компьютерном спорте.

1. Анализ макропоказателей игры.
2. Выявление корреляционных зависимостей.
3. Формулировка причин поражений, составление гипотез.
4. Анализ особенностей игры в защите и нападении, с учетом технико-тактических показателей.
5. Анализ взаимодействия игроков на различных позициях.

Для получения полной картины необходимо провести анализ действий не только своей команды, но и команды соперника.

Однако подобный алгоритм способен иллюстрировать исключительно результаты действий игроков, по итогам соревновательной деятельности. Безусловно получаемые результаты позволяют тренеру сделать определенные выводы и принять соответствующие меры по коррекции слабых сторон подготовленности подопечных.

Таким образом анализ больших данных в компьютерном спорте, это одно из важнейших востребованных средств, необходимых для развития теории спортивной тренировки в компьютерном спорте и подготовки конкурентоспособных спортсменов.

Список источников

1. Гашичев К.М., Сулковский К.Н. Технологии «big data» в профессиональном спорте // Государственное управление: концепции и технологии в эпоху цифровизации : материалы Всероссийской научно-

- практической конференции. СПб. : Изд-во Ленинградского государственного университета им. А.С. Пушкина, 2019. С. 66–71.
2. Моисеев А.Н., Андриянов С.В. Применение методов интеллектуального анализа данных в спорте // Спорт и образование. Перспективы развития УОР в современных условиях : материалы научно-практической конференции. Брянск : Полиграм-Плюс, 2021. С. 130–135.
 3. Лазебная Е.А. Анализ больших данных веб-сервиса поддержки принятия решений в селекционной работе волейбольного клуба // Научные технологии и инновации (XXIII научные чтения) : материалы Международной научно-практической конференции. Белгород : Изд-во Белгородского государственного технологического университета им. В.Г. Шухова, 2019. С. 42–46.
 4. Белоножкина Н.А. Применение цифровых технологий для анализа игровых показателей в баскетболе 3x3 // Физическое воспитание и спорт – взгляд в будущее: интеграция науки и цифровых технологий в образование и практику : материалы Международной научно-практической конференции. М. : ХАССК, 2022. С. 41–45.
 5. CIO // How big data is disrupting the gaming industry. URL: <https://cio.com/article/228327/how-big-data-is-disrupting-the-gaming-industry.html> (дата обращения: 01.06.2022).
 6. Yeung S. Detecting cheaters for multiplayer games: theory, design and implementation // Proc IEEE CCNC. 2006. T. 6. С. 1178–1182.

BIG DATA IN ESPORT

E.A. Kosmina

Lesgaft National State University of Physical Education, Sport and Health, Saint-Petersburg, Russian Federation

Abstract. The article discusses the ways of using big data in the sports industry. The research results are examples of a successful integration of big data analysis into the field of sports, with the possibility of further use of the results in practice. Esports as the most “digital” sports, whose result can be measured in ones and zeros, is the most favorable environment for big data collection and analysis, compared with classical sports. Specific examples of the application of big data analysis methods in esports are given. The system of determining cheaters in first-person shooters, based on the Bayesian network, is considered. Possible directions for big data research in computer sports are outlined. Based on the analysis of literary sources, an algorithm for analyzing big data in esports is proposed.

Keywords: big data, computer sports, esports, data mining, big data, machine learning

ПРИМЕНЕНИЕ УЧЕБНОЙ АНАЛИТИКИ ДЛЯ УПРАВЛЕНИЯ «УСПЕШНОСТЬЮ» СТУДЕНТА

В.В. Крюков, Г.С. Завалин, Д.В. Гмарь, О.А. Батурина, Э.В. Цуранов

Владивостокский государственный университет экономики и сервиса, Владивосток, Россия

Аннотация. В данной работе рассматриваются функциональные возможности сервиса для мониторинга учебной активности студентов, разработанного на платформе Grafana для разных групп пользователей Владивостокского государственного университета экономики и сервиса. Под учебной активностью понимается система показателей, включающая академическую успеваемость, посещаемость занятий, а также активность студентов в электронной образовательной среде. Данный сервис представляет собой систему взаимосвязанных информационных экранов (дашбордов) и позволяет оперативно в интерактивном графическом режиме отслеживать учебную активность студентов в реальном времени с целью выявления обучающихся, склонных к отчислению, а также дисциплин, которые вызывают проблемы с успеваемостью у студентов. В статье также рассматривается предоставление возможности студентам самим отслеживать свои показатели учебной активности в сравнении с средними значениями по группе как элемент дополнительной мотивации обучающихся.

Ключевые слова и словосочетания: учебная аналитика, мониторинг, студенты, учебная активность, дашборд, машинное обучение, успеваемость

Понятие «качество образования» является одним из основных вопросов современной педагогики и общества в целом. Согласно разным точкам зрения, одни исследователи понимают под качеством образования уровень развития личности, другие – качество самого обучения, а некоторые сходятся в том, что взаимосвязь одного и другого. Так, академик РАО М. Поташник определяет качество образования как «соотношение цели и результата, как меру достижения целей, которые заданы операционально и спрогнозированы в зоне потенциального развития обучаемого» [1].

Оценка качества образовательного процесса осуществляется с помощью многогранного мониторинга, который является неотъемлемой частью цифрового обеспечения системы управления учебным процессом университета. Основопологающую роль в этом процессе играет мониторинг учебной деятельности студентов, с помощью которого определяется «успешность» каждого студента в отдельности – то, насколько обучающийся погружен в образовательный процесс на протяжении всего процесса обучения в университете. Основная задача данного инструмента – вовремя выявить студента, имеющего высокую или растущую в динамике вероятность отчисления и оперативно разобраться с причинами возникшей ситуации.

С целью мониторинга учебной активности студентов был разработан инструмент бизнес-анализа «Мониторинг учебной активности и прогнозирование «успешности» студента», имеющий следующие отличительные особенности:

- наличие ролевой модели (студент может просматривать данные своей активности, куратор – учебной группы, студенческий офис – данные всех студентов);
- интерактивное графическое представление данных, с помощью которого можно отслеживать динамику учебной активности в разрезе семестров;
- использование предиктивных моделей для прогнозирования вероятности отчисления студентов по причине неуспеваемости;
- отображение данных в режиме реального времени (как только данные поступают в информационные системы университета, они отображаются в данном сервисе).

Сервис представляет собой набор интерактивных дашбордов (информационных экранов) на платформе Grafana, на котором отображаются различные агрегированные показатели

студентов, представленные в табличном и графическом виде. Общий вид главного дашборда представлен на рис. 1. В верхней части экрана имеются фильтры для выбора интересующей учебной группы, а также текущего семестра, по которым необходимо посмотреть показатели.

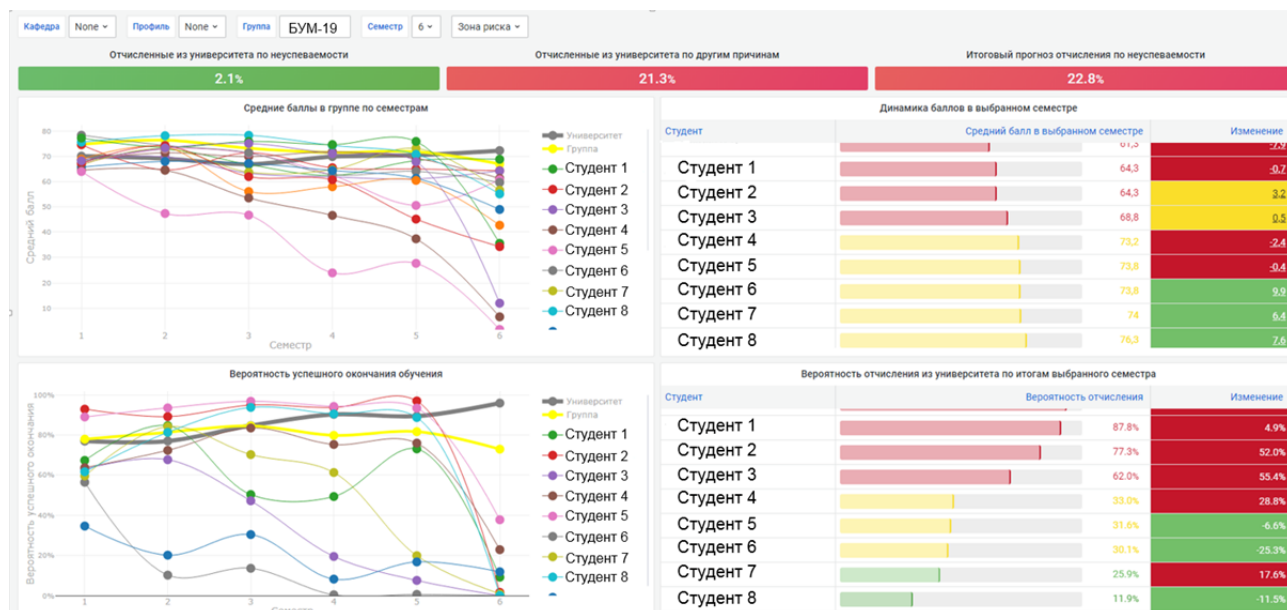


Рис. 1. Общий вид сервиса

Самым простым агрегированным показателем, отражающим уровень академической успеваемости студента, является его средний балл в разрезе по семестрам. Сервис позволяет увидеть динамику изменения среднего балла по каждому из обучающихся в разрезе учебных семестров (рис. 2). С помощью данного графика можно определить в какой момент у обучающегося произошло снижение успеваемости.

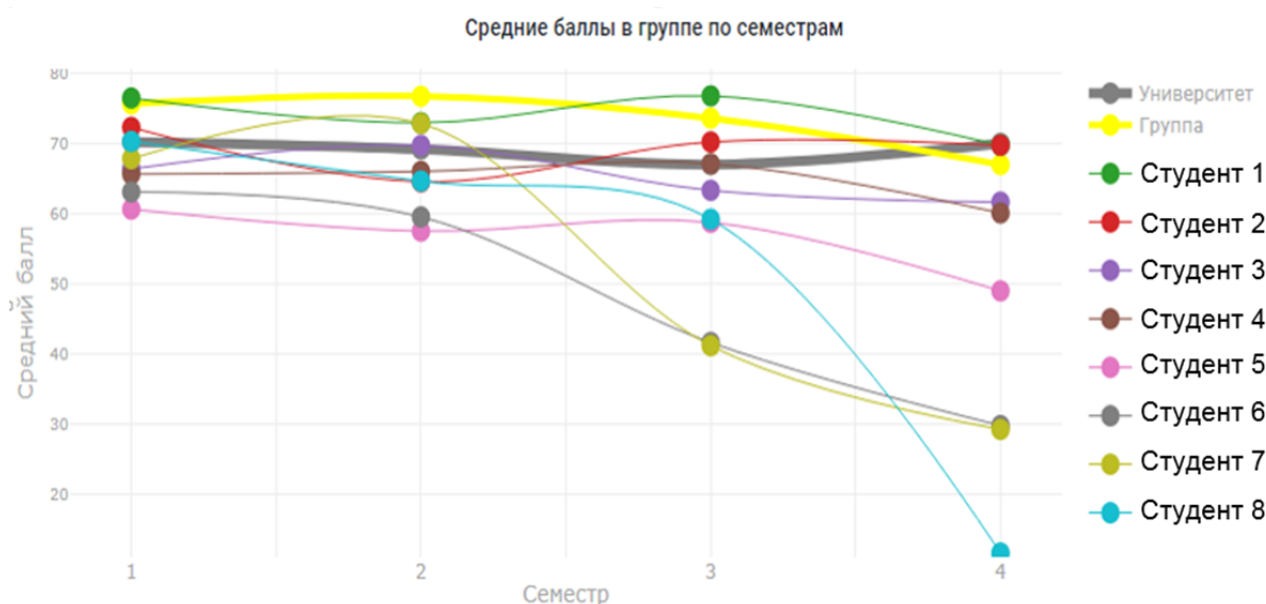


Рис. 2. Динамика среднего балла

На дашборде размещена панель с вероятностью отчисления студентов, на которой для каждого студента предоставляется прогноз вероятности успешно закончить обучение (рис. 3). Расчет вероятности на данном этапе осуществляется по итогам каждого семестра с помощью модели машинного обучения. На панели студенты с высокой вероятностью отчис-

ления выделяются красным – на них стоит в первую очередь обратить внимание кураторам и сотрудникам студенческого офиса.

Модель имеет характер пробного действия: обучена на статических показателях (пол, возраст при поступлении, бюджетная основа или нет) и данных об изменении среднего балла по семестрам, которые были накоплены за предыдущие годы с информацией о том, закончил обучение студент успешно (положительный класс) или же был отчислен по причине неуспеваемости (отрицательный класс). В качестве алгоритма бинарной классификации обучающихся используется логистическая регрессия, которая на указанном наборе признаков на кросс-валидации показывает значения точности (ассигасу), равную 93% (площадь под кривой ошибок – 92%, f1-score – 96%, специфичность – 74%, чувствительность – 95%).



Рис. 3. Вероятность отчисления

В дальнейшем планируется реализовать прогноз вероятности не только по окончании семестра, но и в динамике в процессе обучения. Для этого требуется расширить пул признаков, которые смогут объективно динамично оценивать учебную активность студента по ходу семестра.

В университете реализована система электронных пропусков для прохода на территорию университета, что позволяет по цифровым следам отслеживать посещаемость университета студентами, которая считается как доля посещенных студентом занятий. При этом период посещения должен начинаться не ранее 7 утра, не позднее 17, и заканчиваться не позднее 20, при длительности посещения более 40 минут. Занятие учитывается, если его середина попадает в периоды посещения более чем половиной учебной группы. Точность работы данного алгоритма проверялась с помощью журналов посещаемости учебных групп и оказалась на уровне 80–90% совпадений. В дальнейшем планируется привязать к данному способу учета посещаемости актуальное учебное расписание, чтобы повысить точность показателя.

С помощью перечисленных информационных панелей уже можно выявлять студентов, имеющих значительную вероятность не закончить обучение. Можно погрузиться на уровень ниже: узнать, по каким дисциплинам и в каком семестре у студента имеются академические задолженности. Для этого можно кликнуть на ФИО студента и перейти на дашборд, на котором отображаются показатели учебной активности выбранного студента. Так, можно посмотреть на баллы студента в сравнении со средними результатами по группе (рис. 4). На основании представленной информации можно:

- сделать выводы о том, вызвала ли конкретная дисциплина проблемы у всей группы или же только у одного студента;

- принять решение о необходимости корректировки учебного плана, в случае если дисциплина вызывает проблему у всей группы, а также о целесообразности корректировки методики преподавания с учетом частоты затруднений в данной предметной области у обучающихся.

Успеваемость студента в выбранном семестре

Дисциплина	Балл студента	Средний балл группы
Учебная практика по получению первичн...	0	65.6
Business Evaluation	0	72.4
Финансово-кредитное обеспечение пред...	0	55
Восточный язык базовый (корейский)	0	69.9
Анализ и диагностика финансово-хозяйс...	13	59.4
Бухгалтерский учет модуль 2	15	64.3
Операционный менеджмент	29	59.3
Прикладная физическая культура и спор...	61	82.6
Теория принятия решений	61	65.5
Организационное поведение	62	71.2

Рис. 4. Академические задолженности студента

С целью принятия решений относительно реализуемой дисциплины и ее места в учебном плане образовательной программы был предложен дашборд, на котором отображается сводная статистика по каждой дисциплине (рис. 5). На данном информационном экране можно отследить ряд параметров (как менялся средний балл по этой дисциплине за прошедшие годы, кто ее реализовывал в каждом отдельном семестре, какие технологии применяли, в частности, использовались ли онлайн курсы в рамках изучения дисциплины или ее отдельного модуля, дистанционные технологии или работа в электронной системе и т.д.) как возможные причины снижения успешности освоения.

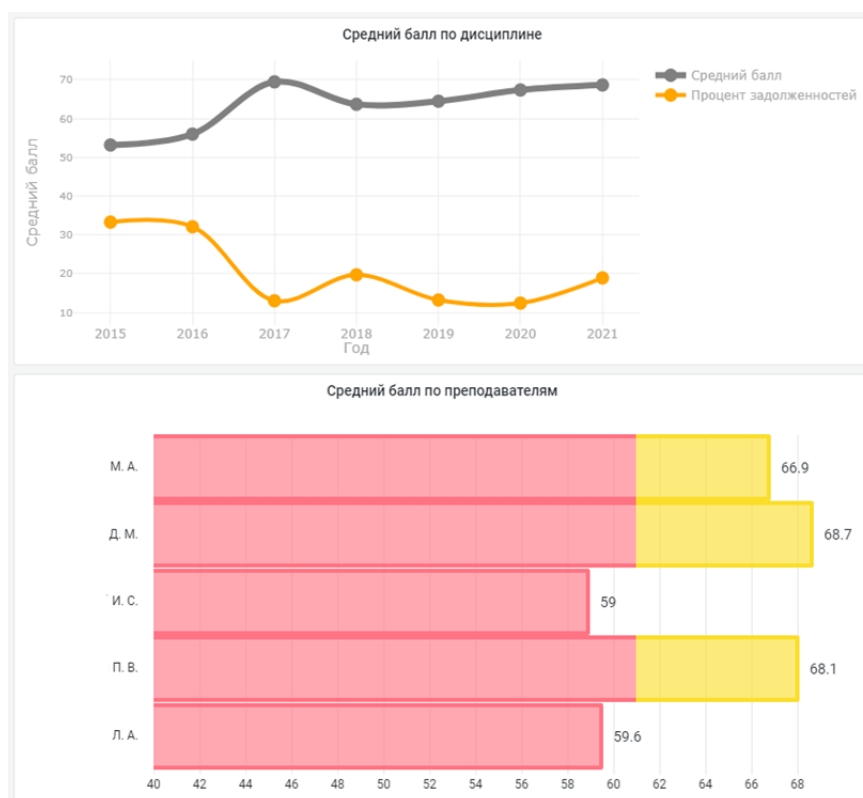


Рис. 5. Статистика по дисциплине

Алгоритм применения сервиса изображен на рис. 6. Сервис позволяет отслеживать учебную активность по семестрам в реальном времени, выявлять студентов с высокой вероятностью отчисления и превентивно принимать меры, способствующие повышению успешности обучающегося. Также с помощью системы взаимосвязанных дашбордов можно определять проблемные дисциплины для студента/группы и корректировать программу обучения, учебный план, методику работы преподавателя.

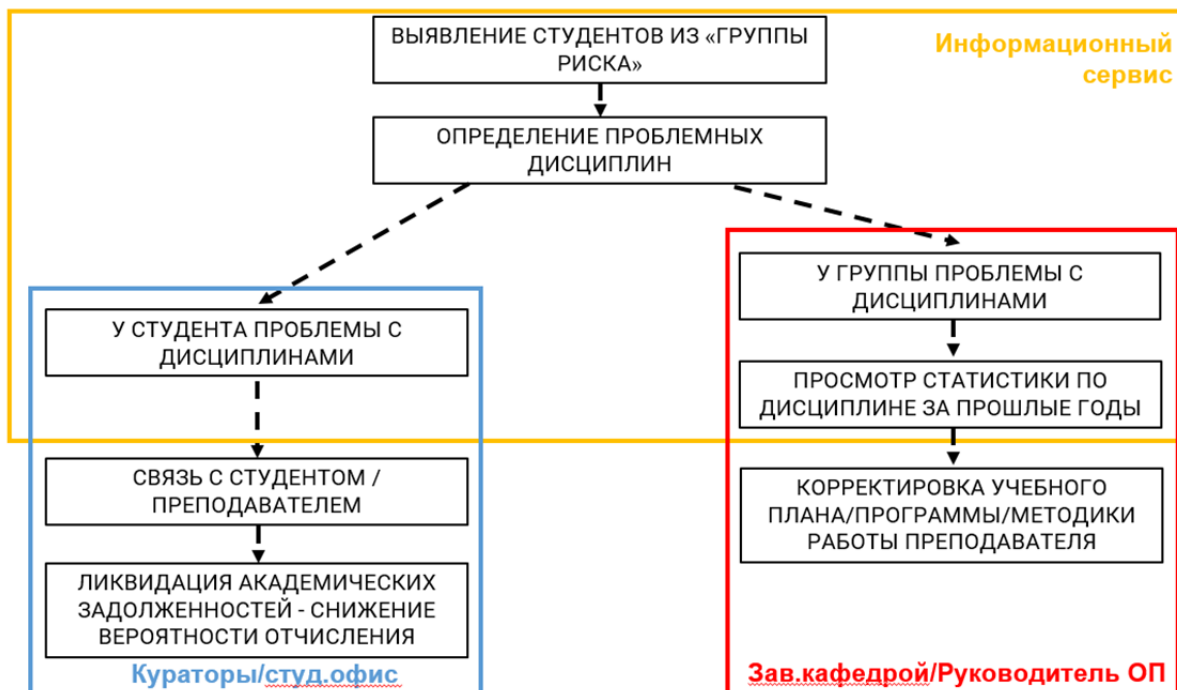


Рис. 6. Алгоритм применения сервиса

Студенты также имеют доступ к данным по своей учебной активности – для них существует отдельный дашборд. Для информирования о текущем состоянии учебной активности используется система светофоров в личном кабинете студента (рис. 7). Зеленый свет горит в случае, если показатели студента находятся на уровне средних по группе или выше, желтый – если немного отстают от них, а красный – если студент по показателям учебной активности существенно отстал от группы. Принцип светофора активно используется специалистами по учебной аналитике в различных учебных заведениях. Например, по данным университета Пердью (Perdue University, Indiana) 89% обучающихся в одном из опросов сочли «сигналы» положительным опытом; 74% заявили, что их мотивация увеличилась с их использованием [2]. Главная цель использования сигналов – сообщить студенту о том, что у него имеются проблемы с учебой, раньше, чем он получит итоговые баллы во время сессии. К тому времени, когда обучающиеся поймут, что у них не все в порядке, может быть уже слишком поздно. Таким образом, «сигналы» направлены на то, чтобы помочь учащимся понять свой прогресс на раннем этапе.

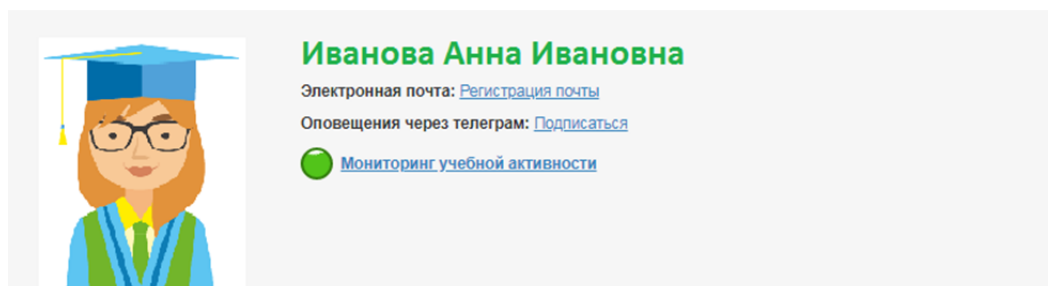


Рис. 7. Светофор учебной активности

Главный «сигнал» у студента завязан на светофорах по каждому из показателей учебной активности (успеваемость, посещаемость, активность в электронной образовательной среде). При нажатии на него студент имеет возможность взглянуть на свои показатели в графическом виде, сравнить, каким образом его показатели отличаются от средних по группе, и сделать соответствующие выводы. На текущем этапе апробации сервиса происходит сбор обратной связи (от студентов), чтобы понять их отношение к сигналам светофора и определить дополнительные показатели учебной активности, которые им хотелось бы видеть на своем персональном дашборде.

Учебная аналитика – это важнейший инструмент, помогающий образовательным организациям решать сложные проблемы, в том числе мотивационные. Основная задача учебной аналитики – помощь студентам: это в том числе и снижение рисков быть отчисленными с помощью повышения образовательных результатов, создание индивидуальных образовательных траекторий. Одной из последних тенденций является доступ самих студентов к результатам учебной аналитики, что может в свою очередь стимулировать их соревноваться со своими сокурсниками и улучшать персональные показатели.

Список источников

1. Поташник М.М. Качество образования: проблемы и технологии управления в вопросах и ответах. М. : Педагогическое общество России, 2002. 352 с.
2. Arnold K.E., Pistilli M.D. Course Signals at Purdue: Using Learning Analytics to Increase Student // Success. Learning Analytics and Knowledge. ACM, 2012.

APPLYING LEARNING ANALYTICS TO MANAGE STUDENT SUCCESS

V.V. Kryukov, O.A. Baturina, D.V. Gmar, G.S. Zavalin, E.V. Tsuranov

Vladivostok State University of Economics and Service, Vladivostok, Russian Federation

Abstract. This article discusses the functionality of the service for monitoring the educational activity of students, developed on the Grafana platform for different groups of users of Vladivostok State University of Economics and Service. Learning activity is understood as a system of indicators, including academic performance, class attendance, as well as student activity in the electronic educational environment. This service is a system of interconnected information screens (dashboards) and allows you to quickly and interactively monitor the learning activity of students in real time in order to identify students who are prone to drop out, as well as disciplines that cause academic performance problems. The article also reviews the ability of students to track their own indicators of learning activity in comparison with the average values for the group as an element of additional motivation for students.

Keywords: learning analytics, monitoring, students, learning activity, dashboard, machine learning, academic performance

ФОРМИРОВАНИЕ КОМПЕТЕНЦИЙ БУДУЩЕГО В УСЛОВИЯХ ОРГАНИЗАЦИОННОЙ НЕОПРЕДЕЛЕННОСТИ

Н.М. Кузнецова

*Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского,
Нижний Новгород, Россия*

Аннотация. Постоянно трансформирующаяся внешняя среда часто выступает сдерживающим для планирования и прогнозирования фактором. По мере расширения своих горизонтов руководитель сталкивается с новым будущим, с одной стороны, полным неопределенности, с другой – новых возможностей и рисков, оказывающих непосредственное влияние на организационное поведение сотрудников. Неопределенность как ситуация неполноты или недостоверности информации оказывает влияние на процессы управления и жизнедеятельности организации. Движущими силами неопределенности могут выступать демографические, социальные, экономические, политические, национальные и технологические факторы. Степень их влияния различна, но так или иначе держит нас в постоянном напряжении, что может негативно отразиться как на внутренних коммуникациях сотрудников в процессе трудовой деятельности, так и на самочувствии общества в целом. Начиная с возникновения управленческой науки, вся энергия исследователей была направлена на минимизацию негативного влияния источников неопределенности на организационное поведение, сопровождающее трудовую деятельность человека. Неопределенность для бизнеса может являться как деструктивным элементом, препятствующим формированию эффективной стратегии становления и развития организации, так и, выступать в роли человеческого познания, идеальным условием для проявления его креативного начала. Акцентирование на требуемых для успешного развития бизнеса целях способствует максимальному получению прибыли для всех заинтересованных участников данного взаимодействия. Основной задачей управленческой деятельности в условиях неопределенности становится формирование нового корпуса сотрудников с высоким уровнем ответственности и мотивации, долей креативности, базирующейся на включенности и вовлеченности в организационные процессы компании, корпуса, способного дать организации возможность выгодно отличиться от других игроков рынка.

Ключевые слова: персонал, неопределенность, толерантность, компетенции будущего, информация, организационное поведение, организационная культура, креативность

Источник финансирования: работа выполнена в рамках реализации программы стратегического академического лидерства «Приоритет–2030».

В условиях нестабильной экономической ситуации, влекущей за собой возникновение для человека различных угроз в отношении его трудовой деятельности, внутренняя структурированная среда компании может выступать безопасным и дружественным местом в это сложное время. Вот почему в основу поиска работы все большее количество соискателей ставят критерий стабильности, обеспечить которую может лишь организация с выстроенной системой ценностей и бережного отношения к персоналу, что, собственно говоря, и закреплено за организационной культурой. Организационное поведение является производной от организационной культуры, в нем находят выражение многие элементы культуры социальной организации (прежде всего, ценности, нормы и традиции).

Существует множество факторов как внешней, так и внутренней среды, влияющих на организационные процессы и поведение персонала в условиях профессиональной деятельности, при этом уровень и степень этого воздействия не всегда заметны. Одним из таких факторов является неопределенность. Большое количество вариаций в определении данного концепта позволяет использовать его во многих направлениях научной деятельности.

R.W. Norton выделил 8 различных категорий, раскрывающих неопределенность: множественность суждений; неточность, неполнота и фрагментированность; вероятность; неструктурированность; дефицит информации; изменчивость; несовместимость и противоречивость; непонятность [1, с. 607–619].

Изучением различных аспектов неопределенности занимаются как зарубежные, так и отечественные авторы: А.Р. MacDonald [2, с. 791–798], S. Bochner [3, p. 393–400],

А.Н. Поддьяков [4, с. 177–194], В.И. Авдийский, В.М. Безденежных [5, с. 46–61], Т.В. Корнилова, М.А. Чумакова [6, с. 92–110], О.Н. Мочалова [7, с. 154–166] и др.

В социологическом поле мы фокусируем свое внимание на процессе выстраивания взаимодействий, формирования коммуникационных сетей между участниками той или иной активности. Неопределенность как ситуация неполноты и недостаточности информации, с одной стороны, влечет за собой определенные риски как в экономическом, так и психологическом плане, с другой – является источником для активных, порой нестандартных действий. Неопределенность сама по себе, выступая отдельным полем изучения, входит в не менее интересное и полиморфное определение креативности. Рассматривая креативность как одну из компетенций будущего, стоит акцентировать внимание на определении ее как целенаправленного процесса по достижению человеком желаемого уровня мотивации, что невозможно без создания среды, позволяющей максимально эффективно использовать личную индивидуальность, талант, компетентность, увлеченность задачей и включенность в процесс, толерантность к неопределенности и прогностическое видение, стремление к развитию и самосовершенствованию.

Организационная культура является важным инструментом создания благоприятной и конструктивной организационной среды, которая отвечает за формирование персонала, обладающего высоким потенциалом, гибкостью, креативностью и адаптивностью. Диагностика организационной культуры позволяет выработать эффективную стратегию маркетинга и менеджмента персонала, поскольку задача методик – получить обратную связь о восприятии организационной культуры в настоящее время и предпочтительной, к которой тяготеет персонал. В зависимости от доминирующего типа культуры руководство может использовать различные инструменты по достижению высокого уровня профессионализма и ключевых навыков своих сотрудников в рамках развития компетенций будущего. Разнообразие данных компетенций, как правило, включает в себя следующие качества успешного специалиста: критическое мышление, эмпатию, гибкость мышления, способность решать трудные задачи, творческое мышление, эффективное управление людьми и проектами, ведение переговоров, умение быстро принимать решения [8] – то, чему не всегда можно научиться в учебном заведении.

Доказательством разворота в сторону мягких навыков служит перечень компетенций для IT-специалистов. Востребованность этой категории неуклонно растет под влиянием международной политической и постпандемийной ситуации, вмещающей им в запрос (наряду с профессиональными качествами) коммуникативность, креативность, самодисциплину, критическое мышление, адаптивность и гибкость [9].

В современном мире документ о высшем образовании перестал выступать гарантом успешного трудоустройства. Цифровая трансформация общества требует от современного человека развития не только профессиональных навыков, получаемых вместе с документом об образовании, но и надпрофессиональных компетенций, ориентированных на умение активно встраиваться в динамичное течение процессов внешней среды. Коронавирусные ограничения бросили обществу вызов, заставив в довольно короткие сроки переформатировать существующую систему взаимодействия. Востребованный и конкурентоспособный специалист будущего должен обладать не только цифровой грамотностью и этикетом, но и набором личностных характеристик, дающих возможность выстраивания индивидуальной траектории профессионального развития, приносящего положительный эффект всему обществу.

Запрос на гибкость, адаптивность, коммуникабельность, стрессоустойчивость, лидерство, креативность не покрывается существующей системой образования. Формирование требуемых компетенций происходит в организациях через имеющуюся систему управления знаниями, что, конечно, не должно снимать ответственности с вузов в части достижения заданных компетенций. В процессе обучения персонала компании ориентируются на компетентностно-ориентированный подход, вызванный к жизни социально-экономическими предпосылками, когда бизнес и рынок предъявляют к специалистам новые требования, направленные на достижение максимальной продуктивности в конкретном виде деятельности. Подключение обра-

зовательных учреждений к данной системе, коллаборация бизнеса и образования позволят выстроить эффективную систему управления знаниями и человеческим капиталом.

Компетенция как знание, умение и опыт помогает справиться с большим количеством профессиональных задач. Компетентность как качественная характеристика предполагает наличие фактора внутренней мотивации к дальнейшему развитию и эффективному осуществлению профессиональной деятельности. Организационная среда компании как инструмент формирования человеческого капитала с его разнообразными компетенциями представляет для исследователей большой интерес. Через ее диагностику мы можем понять, насколько она способна дать человеку ключ к его успешной интеграции в социально-экономическую систему.

Как правило, инструментом организационной диагностики выступает методика диагностики организационной культуры – методика (OCAI – Инструменты оценки организационной культуры) К. Кэмерона и Р. Куинна [10]. Вузы сочетают в себе несколько видов культур, как правило, базируясь на бюрократическом типе [11].

Бюрократизация организационной среды не способна предоставить возможности для развития творческих и самостоятельных начал, что еще больше увеличивает разрыв между образовательной средой и запросами современного рынка труда. Как правило, формальный менеджмент и бюрократическая структура больше подходят для выполнения государственных заданий с четко прописанными сроками, формами отчетности и процедурами. Создание нового и уникального продукта, услуги, программы, требует творческого подхода, самоорганизации, желания и заинтересованности в процессе. Поэтому несмотря на имеющиеся в вузе элементы иерархии, подтверждением понимания происходящих процессов в обществе служит смещение акцентов в желаемой организационной культуре вузов в сторону новаций и конкурентной среды.

Организационная культура некоторых современных компаний носит клановый характер. Диагностика организационной культуры одного из нижегородских кинотеатров выявила следующие тенденции в части восприятия и предпочтения культуры сотрудниками компании (n = 15) (рис. 1).

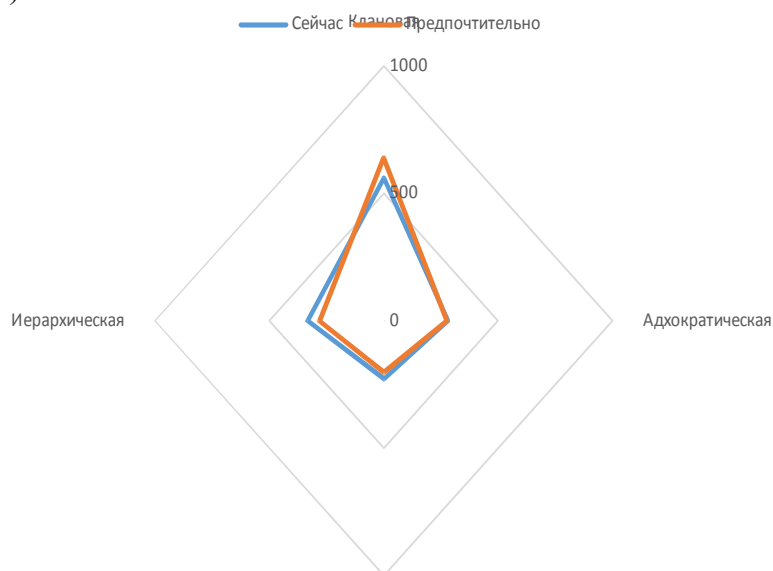


Рис. 1. Организационная культура организации АО «Кино»

Сотрудники воспринимают и предпочитают оставаться в клановом типе организационной культуры, причем вектор движения в этом направлении прослеживается достаточно четко. Пандемия и международная политическая ситуация нанесли серьезный удар по отрасли киноиндустрии. Возможно, поэтому представленные результаты выглядят таким образом. Клановый тип культуры базируется на коллективном виде работы, больше напоминая семью,

чем коммерческую структуру. Компания ориентирована на формирование комфортной внутренней среды, заботясь о профессиональном развитии персонала, поддерживая высокий уровень лояльности к месту осуществления трудовой деятельности. Организация с данным типом культуры закрыта, ориентирована на внутреннюю среду и способна сформировать лишь часть требуемых для эффективной деятельности компетенций (в частности, командную работу, коммуникативные навыки, эмпатию), но этого недостаточно для способности противостоять постоянно меняющимся условиям внешней среды. Отсутствие конкуренции в коллективе, опыта индивидуального взаимодействия с иными профессиональными группами и как следствие – недостаточность квалификации в области принятия решений и действий в условиях риска не дают развивать весь спектр надпрофессиональных навыков. Для организаций, находящихся в клановом типе культуры, большая опасность таится в неподготовленности сотрудников к ответной реакции на трансформацию внешней среды как в экономическом, политическом, так и в социальном плане, поскольку любые изменения воспринимаются достаточно враждебно и тяжело. Желание сотрудников находиться в клановом типе организационной культуры продиктовано стремлением чувствовать себя нужным, значимым, частью целого в поисках безопасной и комфортной среды в данной конкретной компании.

Выступая самым важным ресурсом в современном цифровом обществе, дефицит информации накладывает отпечаток на восприятие организационной культуры, меняя организационное поведение сотрудников и траекторию формирования компетенций. Важно проследить, насколько данная зависимость меняет парадигму развития преподавательского кластера, выступающего источником формирования системы знаний. Современный выпускник должен уметь адаптироваться под разные организационные культуры, при этом в любой из них ощущая себя эффективным и продуктивным. Существующая взаимосвязь организационной неопределенности с воспринимаемыми и предпочитаемыми типами организационной культуры среди преподавательского состава вуза ($n = 73$) имеет следующие показатели (табл. 1): чем выше показатели организационной неопределенности по всем представленным категориям, тем меньше возможностей для организации сформировать клановый тип организационной культуры, что не представляет возможности разворота в сторону личности, поддержания развития и поступательного профессионального развития. Чем выше показатели организационной неопределенности межличностного и трудового характера, тем больше стремление персонала компании уйти в сторону иерархического типа культуры с регламентированными законами, правилами и коммуникациями, которая не дает возможности эффективного и активного развития, побуждая сотрудников к действиям, заведомо прописанным и регламентированным, лишенным инициативы и креатива. Таким образом, мы можем отметить, что неопределенность влияет на восприятие организационной культуры, типы которой определяют модели поведения и взаимодействия персонала в компании.

Создавая непонятную для сотрудника ситуацию на рабочем месте, неопределенность выступает дезориентирующим, раздражающим и пугающим фактором, источником стресса и негатива. Последствия действий в условиях неопределенности непредсказуемы, поскольку дефицит информации о факторах внутренней и внешней среды искажает процессы как управленческой деятельности, принятия решений руководящего состава, так и реализации поставленных задач перед сотрудниками образовательных учреждений, несущих в себе важные для формирования социально-активного, образованного, ответственного, патриотичного гражданина, полноценного члена современного общества.

Руководству организации следует сфокусироваться на разработке программ для снижения уровня неопределенности и ее трансформации в информированную реальность и перспективу. Неопределенность побуждает человека к активным действиям, поскольку размывает границы возможностей и перспектив. Персонал, имеющий высокую толерантность к неопределенности, обладает повышенными способностями в решении новых, сложных, творческих задач, где последствия решений не всегда предсказуемы, и обладает лучшими коммуникативными свойствами. Целью образовательных учреждений является конструирование

организационной среды, с одной стороны, базирующейся на принципах, ценностях, нормах и правовом поле современного социума, с другой – ориентированной на помощь в достижении человеком профессиональной, личной успешности, а также на признание его обществом.

Взаимосвязь организационной неопределенности с воспринимаемыми в организации и предпочитаемыми типами организационной культуры у профессорско-преподавательского состава ННГУ

Показатель	Неопределенность в отношении организации	Управленческая неопределенность	Межличностная неопределенность	Неопределенность в отношении непосредственного руководителя	Трудовая неопределенность	Организационная неопределенность в целом
Восприятие организационной культуры организации						
Как клановой	-0,38	-0,52	-0,31	-0,34	-0,35	-0,45
Как адхократической	-0,14	-0,19	-0,19	-0,10	-0,20	-0,20
Как рыночной	-0,15	0,02	-0,19	-0,03	-0,17	-0,11
Как бюрократической	0,18	0,23	0,32	0,07	0,35	0,27
Предпочтение организационной культуры						
Клановой	-0,09	0,14	-0,13	-0,08	-0,08	-0,05
Адхократической	-0,04	-0,15	-0,03	0,00	-0,07	-0,07
рыночной	0,13	-0,01	0,23	0,04	0,23	0,14
Бюрократической	0,17	0,10	0,35	0,02	0,23	0,20

Примечание. Достоверные ($p < 0,05$) корреляции выделены штриховкой ячеек.

В рамках управления человеческими ресурсами организации происходит смещение акцента в сторону создания устойчивости в мире изменений. Формирование нового корпуса персонала, толерантного к различного рода колебаниям, путем поощрения сопричастности, постоянного развития, повышения профессиональной ценности каждого работника через максимизирование чувства вклада в достижение общих значимых целей – одна из парадигм управления неопределенностью и выстраивания компетенций будущего.

В таких условиях адаптация, самоактуализация, развитие креативности и активности действий, гибкости и самостоятельности, формирование толерантности к неопределенности происходит значительно быстрее, что позволяет остаться в плюсе как обществу, так и отдельной личности, через обретение уверенности и устойчивости в процессе происходящих изменений в VUCA мире.

Список источников

1. Norton R.W. Measure of ambiguity tolerance // Journal of Personality Assessment. 1975. Vol. 39 (6). P. 607–619.
2. MacDonald, Alan P. Revised scale for ambiguity tolerance: reliability and validity // Psychological Reports. 1970. Vol. 26. P. 791–798. URL: <https://www.sgha.net/library/pr0%252E1970%252E26%252E3%252E791.pdf> (дата обращения: 01.06.2022).
3. Bochner S. Defining Intolerance of Ambiguity // The Psychological Record. 1965. Vol. 15. P. 393–400. URL: <https://link.springer.com/journal/40732/volumes-and-issues/15-3> (дата обращения: 30.05.2022).
4. Подьяков А.Н. и [др.]. Неопределенность в решении комплексных проблем // Человек в ситуации неопределенности. М. : Теис, 2007. С. 177–194. URL: <http://geum.ru/next/art-268904.php> (дата обращения: 13.06.2022).
5. Авдийский В.И., Безденежных В.М. Неопределенность, изменчивость и противоречивость в задачах анализа рисков поведения экономических систем // Эффективное антикризисное управление. СПб., 2011. URL: <https://www.elibrary.ru/item.asp?id=16753895> (дата обращения: 31.05.2022).
6. Корнилова Т.В., Чумакова М.А. Шкалы толерантности и интолерантности к неопределенности в модификации опросника С. Баднера // Экспериментальная психология. 2014. № 1. С. 92–110.
7. Мочалова О.Н., Ромашкин С.В. и др. Мотивация достижения как фактор конформного реагирования в ситуации неопределенности // Человек в ситуации неопределенности. М. : ТЕИС, 2007. С. 154–165.
8. Навыки будущего: какие качества будут больше всего цениться на рынке труда в 2022 году. URL: <https://businessman.ru/post/navyiki-buduschego-kakie-kachestva-budut-bolshe-vsego-tsennitsya-na-ryinke-truda-v-godu.html?> (дата обращения: 10.06.2022).

9. Архипова Ю. Soft skills для разработчиков: зачем и как развивать? URL: <https://habr.com/ru/post/509278/> (дата обращения: 11.06.2022).
10. Камерон К., Куинн Р. Диагностика организационной культуры. СПб., 2001. 320 с.
11. Кузнецова Н.М. Организационная неопределенность как фактор восприятия организационной культуры вуза // Социология. 2020. № 4. С. 205–214. URL: <https://www.elibrary.ru/item.asp?id=43989344> (дата обращения: 30.05.2022).

FORMING COMPETENCIES OF THE FUTURE IN CONDITIONS OF ORGANIZATIONAL UNCERTAINTY

N.M. Kuznetsova

*Lobachevsky State University of Nizhny Novgorod – National Research University,
Nizhny Novgorod, Russian Federation*

Abstract. The constantly transforming external environment often acts as a deterrent for planning and forecasting. As their horizons expand, leaders are faced with a new future full of uncertainty, on the one hand, and new opportunities and risks that have a direct impact on the employees' organizational behavior, on the other. Uncertainty as a situation of incompleteness or unreliability of information affects the management processes and the life of the organization. Uncertainty can be driven by demographic, social, economic, political, national and technological factors. The degree of their influence is different, but one way or another keeps us in constant tension, which can negatively affect both the internal communications of employees in the process of work, and the well-being of society as a whole. Since the emergence of management science, all the energy of researchers has been aimed at minimizing the negative impact of sources of uncertainty on organizational behavior that accompanies a person's labor activity. For business, uncertainty can be both a destructive element that hinders the formation of an effective strategy for the formation and development of an organization, and, acting as human knowledge, an ideal condition for the manifestation of its creative beginning. Focusing on the goals required for successful business development contributes to the maximum profit for all interested participants in this interaction. The main task of managerial activity in conditions of uncertainty is the formation of a new corps of employees, with a high level of responsibility and motivation, a share of creativity based on inclusion and involvement in the organizational processes of the company, capable of giving the organization the opportunity to distinguish itself favorably from other market players.

Keywords: personnel, uncertainty, tolerance, competencies of the future, information, organizational behavior, organizational culture, creativity

СЦИЕНТИЗАЦИЯ БЫТА В РОССИЙСКИХ МЕГАПОЛИСАХ: ЭТАПЫ ВНЕДРЕНИЯ ИННОВАЦИЙ С 2000 ПО 2022 г.

Ю.В. Лагутин

*Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского,
Нижний Новгород, Россия*

Научный руководитель: И.Э. Петрова, д-р социол. наук, доцент

Процесс сциентизации быта жителей современного мегаполиса многогранен. Для проведения его изучения необходима систематизация подходов, проработка внедрения инноваций в сфере науки и технологий. На основе проведенной систематизации процессов сциентизации быта в современном мегаполисе и подходов к его изучению становится возможным определить основные этапы внедрения в повседневную жизнь городских жителей передовых научных достижений в форме прикладных технологий, влияющих на индивидуальное восприятие типичных проблем, связанных с взаимодействием с другими людьми в пространстве мегаполиса. Иногда это влияние может показаться менее очевидным, чем есть на самом деле, а его интеграция в быт современного общества зачастую уже воспринимается как данность. При этом разделение и частичная диффузия личного и общественного пространства становится ключевой темой при рассмотрении тенденций технологизации быта последних двух десятилетий.

Ключевые слова: сциентизация, быт, инновации, урбанизация, систематизация, мегаполис, социальные практики, общество

В качестве отправной точки для проведения исследования выбран рубеж 1990–2000-х гг. сразу по нескольким причинам.

Во-первых, информационные технологии, ориентированные на коммерческое использование интернета, получают в этот период значительную поддержку инвесторов и признание со стороны финансового сектора, оценивших в развитии и внедрении технологических инноваций возможность появления новых рынков и моделей потребительского поведения. Несмотря на последующую нормализацию бизнес-среды после так называемого «бума дот-комов» (ажиотажа на финансовых рынках, вызванных значительной переоценкой акций компаний, ориентированных на бизнес-модели с использованием интернета и зарегистрированных в домене «.com»), инновации, ориентированные на жителя современного мегаполиса, доказали свою востребованность и жизнеспособность [1].

Во-вторых, в массовой культуре на рубеже 1990–2000-х гг. распространяются (не в последнюю очередь из-за внедрения новых технологий 3D-моделирования и цифровой обработки изображения) образы будущего, связанного с использованием в рамках повседневного взаимодействия технологий виртуальной реальности, искусственного интеллекта, «умных» помощников (роботов и микроботов) и устройств, частичной киборгизации, а также опосредованного с помощью информационно-телекоммуникационных сетей присутствия в любой точке земного шара (и космоса). В связи с данной ролью кинематографа как социального визионера можно упомянуть такие фильмы, как «Матрица», «Экзистенция», «Тринадцатый этаж» и другие, позволившие донести до широких масс идеи из философских трактатов и фантастической литературы [2, 3].

В-третьих, распространение интернета и информационных технологий в значительной степени диверсифицировало источники социально значимой информации, подорвав монополию телевидения и печатной прессы, что позволяет говорить о новой конфигурации сил в гражданско-политической сфере. Национальные государства в этот период делают первые шаги в сфере регулирования деятельности транснациональных корпораций, в том числе – в сфере информационных технологий, а также разрабатывают первые варианты мероприятий, связанных с целенаправленной политикой по развитию электронного взаимодействия власти, бизнеса и граждан. Подобные действия свидетельствовали, прежде всего, о том, что интер-

нет и информационные технологии прошли стадию «игрушки для гиков» и превратились в объект интереса со стороны органов власти вследствие своего широкого распространения и потенциальной интеграции в повседневные социальные практики [4].

Наконец, нельзя не упомянуть «Проблему 2000» – комплекс массовых полуапокалиптических ожиданий от неготовности и/или неспособности компьютеров, запрограммированных на восприятие обозначения года в формате двух последних цифр, справиться с переходом от 1999 г. (две последние цифры «99») к 2000 г. (две последние цифры «00», как и в 1900 г.), что могло бы вызвать сбой всех электронных систем и потенциальное отключение инфраструктуры жизнеобеспечения в городах. Данная проблема, активно обсуждаемая в массовом формате в течение всей второй половины 1999 г., позволила перенести в центр мировой повестки вопросы влияния компьютерных технологий и инициировавших их появление научных исследований на устойчивость и качество жизни обычного человека [5].

Как было отмечено выше, основные этапы сциентизации и технологизации повседневных практик жителей современного города разворачивались в пространстве индивидуально-личного потребления (личное пространство дома, автомобиля, рабочего места) и в рамках устойчивых социальных практик и вариантов социального взаимодействия (общественные пространства и объекты, в том числе – виртуальные), а также на границах этих пространств и в зонах их взаимопроникновения и смешения. В связи с этим отдельные элементы сциентизированного бытового окружения получают разное воплощение и применение в личном и общественном пространстве в зависимости от рассматриваемого этапа.

В обобщенном виде рассматриваемые ниже этапы внедрения и сценарии использования элементов технологизации повседневной жизни можно представить в виде таблицы.

Хронология и характеристика этапов внедрения научно-технологических инноваций в жизни современного мегаполиса

Название и временные границы этапа	Характеристика этапов		
	Личное пространство/ индивидуальное потребление	Общественные пространства/ социальное взаимодействие	Сочетание личного и общественного пространства / личного потребления и социального взаимодействия
1. Развитие коммуникационных возможностей компьютерной техники (2000–2007)	Эффективный поиск информации об объектах «реального» мира с целью покупки необходимых товаров и услуг /	Развитие флешмобов и технологий умных толп при координации массовых акций и мероприятий /	Возникновение агрегаторов онлайн-объявлений о продаваемых товарах, оказываемых услугах /
2. Внедрение мобильных технологий и потребления контента (2002–2009)	Использование интернета для поиска книг, фильмов и другой информации в цифровой форме	Развитие социальных сетей для поиска наиболее подходящего социального окружения	Развитие использование сервисов поиска работников и потенциальных партнеров для личных отношений
3. Формирование возможностей для творчества и обмена его результатами с помощью высокотехнологичных устройств (2007–2015)	Использование индивидуальных устройств для творческого самовыражения (запись видео, фотографирование)	Краудсорсинг как способ создания информационных ресурсов и решения социальных проблем	Поддержка наиболее талантливых авторов с помощью специальных платформ, развитие благотворительных платформ /
4. Распространение «умных помощников» (smart gadgets) и «носимой электроники» (wearable electronics) (2011–2017)	Формирование индустрии «отслеживания» параметров здоровья, потребления, финансового поведения, использование отдельных элементов «умного дома»	Развитие «мобильной картографии» и картографических сервисов, позволяющих корректировать социально значимую информацию	Обмен данными и геймификация индивидуальных и социальных практик, координация усилий отдельных потребителей ограниченных ресурсов

Название и временные границы этапа	Характеристика этапов		
	Личное пространство/ индивидуальное потребление	Общественные пространства/социальное взаимодействие	Сочетание личного и общественного про- странства / личного по- требления и социального взаимодействия
5. Развитие дистанционных технологий в сфере работы, учебы и других форм социального взаимодействия (2013–2019)	Перенос части работы в индивидуальное пространство, развитие сервисов доставки и профессиональных услуг	Развитие электронных государственных услуг в рамках разрешения индивидуальных жизненных ситуаций, развитие электронных ассистентов и помощников в рамках социального взаимодействия	Формирование среды обмена рабочими операциями, переосмысление образовательной среды в контексте развития «эдьютей-нмента» (как сочетания образования и развлечения)
6. Генерирование и обработка «больших данных», развитие инфраструктуры «умного города» (2016–2022)	Вовлечение граждан в полноценное цифровое взаимодействие с корпоративной сферой и с органами власти посредством предоставления персональных данных и сведений о социальных микротранзакциях в обмен на высокий уровень индивидуализации обслуживания и потребления	Развитие подходов в рамках координации социальных действий и построения предиктивных моделей социального поведения, использование геоинформационных систем в рамках низовых гражданских инициатив	Развитие моделей экономики совместного пользования жилья, автомобилей, одежды и других ресурсов, а также индивидуализированных форматов массового производства и потребления

Источник: разработано автором.

Отметим, что отдельные этапы возникают на фоне предшествующих и в фоновом режиме продолжают актуализировать отдельные тенденции сциентизации быта после зарождения и развития последующих стадий научно-технологического развития в рассматриваемой сфере. В качестве временных рамок указаны годы наибольшей активности в рамках того или иного этапа и восприятия.

Этап 1. Развитие коммуникационных возможностей компьютерной техники (2000–2007 гг.). Данный период характеризуется переходом к восприятию компьютера как точки входа в интернет с его возможностями безграничного взаимодействия и поиска необходимой информации научного, развлекательного и бытового характера. Переход к HTML 4 (от англ. Hypertext markup language – «язык гипертекстовой разметки»), дальнейшее развитие интернет-браузеров (Internet Explorer, Netscape, Opera) и поисковых систем (Яндекс, Yahoo, Google, Baidu) позволили наряду с развитием аппаратной части – появлением полноценных портативных компьютеров, КПК (карманных персональных компьютеров, PDA – Personal digital assistant), мобильных телефонов с возможностью выхода в интернет – обеспечить невиданные ранее коммуникационные возможности жителю большого города, сделав его повседневную жизнь более насыщенной и полной [6, 7]. Дальнейшее развитие в рамках данного этапа существовавших ранее сайтов онлайн-торговли Ebay и Amazon (в России – Ozon.ru), сервиса объявлений Graigslist, мессенджеров IRC и ICQ (наряду с возникшим позднее Skype) позволило жителям мегаполиса почувствовать себя обитателями «глобальной деревни», где все если не знают друг друга, то, по крайней мере имеют возможность пообщаться, продать, купить или назначить встречу. Исходя из этого, наиболее репрезентативными в рамках данного этапа являлись следующие сценарии внедрения и использования научно-технологических инноваций в жизни российского мегаполиса, выделяемые на основе преобладающего формата их использования (индивидуального, коллективного или смешанного).

Этап 2. Внедрение мобильных технологий и потребления контента (2002–2009 гг.). Указанный период характеризуется, прежде всего, дальнейшим развитием мобильных телефонов с целью потребления развлекательного и иного контента, а также диверсификации социальных контактов. Появление первого удобного для использования бесклавиатурного мобильного те-

лефона от компании Apple, интерфейс которого был ориентирован на подключение к интернету, способствовал усилению степени персонализации использования устройств этого класса и их интеграции в повседневную жизнь жителей мегаполиса. Запуск в 2008 г. онлайн-магазина приложений для мобильного телефона в свою очередь обеспечил реализацию поведенческих паттернов и пользовательских схем в рамках конкретных жизненно-бытовых ситуаций: планирование личного времени, прослушивание музыки, покупка товаров, заказ еды [8]. Отдельного упоминания заслуживает процесс поиска и загрузки аудио- и видеопроизведений в интернете (зачастую нелегально) и развития пиринговых сетей для обмена файлами. При этом упомянутая выше ориентация на потребление контента и социальное взаимодействие по этому поводу становится ключевой темой рассматриваемого периода сциентизации и технологизации повседневных практик горожан, реализующих следующие сценарии.

Этап 3. Формирование возможностей для творчества и обмена его результатами с помощью высокотехнологичных устройств (2007–2015 гг.). Импульс для развития и широкого распространения, полученный социальными сетями и сервисами для обмена сообщениями (Twitter) в предшествующий период, в рамках данного этапа получает развитие в рамках новых сценариев внедрения научно-технологических инноваций в повседневную жизнь горожан.

Основным лейтмотивом этапа 3 сциентизации повседневных практик жителей мегаполиса является дальнейшее углубление интерактивности и появление полноценных виртуальных общественных пространств, обеспечивающих самовыражение участников и решение реальных социальных проблем.

Этап 4. Распространение «умных помощников» (smart gadgets) и «носимой электроники» (wearable electronics) (2011–2017 гг.). Рассматриваемые в данном периоде устройства, содержащие результаты научных разработок и воспринимаемые как потребительские инновации в пространстве мегаполиса и в разворачивающихся в нем социальных практиках, нашли отражение в концепции «умного окружения» или, в бытовом контексте, в элементах и решениях «умного дома» (smart house). В связи с этим отметим наиболее устойчивые сценарии внедрения и использования научно-технологических инноваций на социально-бытовом уровне [9].

Этап 5. Развитие дистанционных технологий в сфере работы, учебы и других форм социального взаимодействия (2013–2019 гг.). Этот период связан с развитием зародившихся ранее тенденций, связанных с оптимизацией личных, корпоративных и государственных расходов путем перехода на дистанционные форматы работы и социального взаимодействия [10]. Личное присутствие и общение постепенно становится своеобразной роскошью, а не досадной необходимостью, в связи с чем можно выявить следующие сценарии внедрения и использования инноваций в повседневной городской жизни.

Этап 6. Генерирование и обработка «больших данных», развитие инфраструктуры «умного города» (2016–2020 гг.). Предшествующее развитие новых технологий и инновационных решений в пространстве города после внедрения ряда аналитических и интеграционных механизмов позволили перейти к следующему этапу, характеризующемуся постепенным не всегда явным переходом от фиксации и анализа социальных транзакций отдельных индивидов и общностей к их моделированию в нужном ключе, руководствуясь коммерческими интересами, требованиями безопасности или требованием максимизации общественных благ для каждого индивида в процессе их распределения и использования. Следует отметить и изменение экологической повестки в сторону личной ответственности каждого жителя мегаполиса вместо обвинений безличной «империи зла» транснациональных корпораций.

Закономерным выводом по итогам рассматриваемого этапа будет служить признание слома традиционных отношений жителей большого города к вещам, находящимся в их владении или пользовании, а также постепенный переход к минималистичным и устойчивым с точки зрения воздействия на экологические системы социальным практикам.

Анализ этапов внедрения научно-технологических инноваций в повседневные социальные практики жителей современного мегаполиса позволяет определить следующие общие для всех рассмотренных элементов моменты и тенденции.

А. Сфера сочетания индивидуальных сценариев и коллективных форматов взаимодействия жителей городов в процессе использования новых технологических решений является источником трендов и идей для последующих этапов рассматриваемого процесса сциентизации быта в современном мегаполисе.

Б. В качестве экспериментальных площадок и источников инноваций можно рассматривать выставки потребительской электроники, фестивали гик-культуры и форсайт-сессии на корпоративном уровне и с участием представителей власти.

В. На протяжении рассматриваемого периода (2000–2020 гг.) роль заказчика инноваций постепенно переходит от венчурных фондов и национальных инновационных систем к потенциальным пользователям, что позволяет говорить об изменении баланса интересов в развитии данной сферы.

Наконец, необходимо отметить, что само инновационное развитие и процесс внедрения продуктов научно-технологического решения в повседневную жизнь современных горожан происходят скачкообразно, на основе порывов, вызванных временным энтузиазмом, и откатов, обусловленных частичным разочарованием в результатах. Указанное обстоятельство требует рассмотрения заявленной темы в контексте сочетания различных индивидуальных и коллективных интересов.

Список источников

1. Толкачев С.А., Попов А.К. Формирование финансовых пузырей на стадии роста экономической системы // Вестник Финансового университета. 2015. № 2. С. 84–95.
2. Прими красную таблетку: наука, философия и религия в «Матрице» / пер. с англ.; под ред. Г. Йефетта. М. : Ультра. Культура, 2003.
3. Миронов В.В., Сокулер З.А. Тоска по истинному бытию в цифровой культуре // Вестник Московского университета. Серия 7. Философия. 2018. № 1. С. 3–22.
4. Кузнецов С. Ощупывая слона: Заметки по истории русского Интернета. М. : Новое литературное обозрение, 2004.
5. Фрунзе А. «Проблема 2000» и ваш путь к успеху // Компоненты и технологии. 2000. № 2. С. 3–4.
6. Малсид М, Вайз Д. Google. Прорыв в духе времени / пер. с англ. М. : Эксмо, 2007.
7. Арутюнов В.В. О современных мобильных компьютерных платформах // Вестник Московского финансово-юридического университета. 2013. № 3. С. 172–178.
8. Айзексон У. Инноваторы. Как несколько гениев, хакеров и гиков совершили цифровую революцию / пер. с англ. М. : Corpus, 2015.
9. Гринфилд А. Радикальные технологии: устройство повседневной жизни / пер. с англ. И. Кушнаревой. М. : Дело, 2018. С. 238–240.
10. Срничек Н., Уильямс А. Изобретая будущее. Посткапитализм и мир без труда / пер. с англ. Н. Охотина. М. : Strelka Press, 2019.

SCIENTIZATION OF EVERYDAY LIFE IN RUSSIAN METROPOLISES: STAGES OF INNOVATION IMPLEMENTATION FROM 2000 TO 2022

Yu.V. Lagutin

*Lobachevsky State University of Nizhny Novgorod – National Research University,
Nizhny Novgorod, Russian Federation*

Abstract. The process of scientization of the life of the inhabitants of a modern metropolis is multifaceted. To conduct its study, it is necessary to systematize approaches and study the introduction of innovations in the field of science and technology. Based on the systematization of the processes of everyday life in a modern metropolis and approaches to its study, it becomes possible to determine the main stages of the introduction into the daily life of urban residents of advanced scientific achievements in the form of applied technologies that affect the individual perception of typical problems associated with interaction with other people in the space of a metropolis. Sometimes this influence may seem less obvious than it actually is, and its integration into the everyday life of modern society is often already perceived as a given. At the same time, the separation and partial diffusion of personal and public space becomes a key topic when considering the trends in the technologization of everyday life of the last two decades.

Keywords: scientization, everyday life, innovations, urbanization, systematization, megapolis, social practices, society

АНАЛИЗ ДИСКУРСОВ БЛАГОТВОРИТЕЛЬНОСТИ ЧЕРЕЗ ПРИЗМУ БИБЛИОМЕТРИЧЕСКОЙ СЕТИ

Е.С. Мищенко, А.А. Барышев, А.В. Губанов

Национальный исследовательский Томский государственный университет, Томск, Россия

Феномен благотворительности во все эпохи отражает как уровень социально-экономического развития общества, так и состояние нравственной культуры социума в целом, его цивилизационную зрелость. В данной работе описывается эксперимент по картированию дискурса, формируемого свежими статьями, содержащими лингвистические маркеры благотворительности: «Charity», «Philanthropy», «Giving», «Fundraising», «Non-profit» и др. Целью исследования является определение возможности выявления тематической дифференциации дискурса благотворительности на основе построения сети слов, употребляемых совместно с названными маркерами. Исходную базу данных составили названия и аннотации, относящиеся к более чем 5 000 статей 2020–2022 гг. публикации, найденных по указанным маркерам в базах Web of Science (WoS) и Scopus. В качестве метода автоматической обработки текстов был применен метод сетевого анализа совместной встречаемости слов, что позволило получить следующие результаты. Во-первых, построена сеть слов и словосочетаний на основе коэффициента Жаккара: чем выше частота совместного употребления слов по отношению к общему количеству анализируемых текстов, в которых встречается хотя бы одно из них, тем больший вес получают эти слова, наглядно увеличиваясь на карте. Во-вторых, выявлены кластеры, относящиеся к различным тематическим областям, трактующие различные аспекты современной благотворительности. Данные результаты могут быть полезны для исследования тематического разнообразия проявлений благотворительности в более значительных временных масштабах, а также для совершенствования методологии дифференциации дискурсов по гомогенным тематическим кластерам.

Ключевые слова: благотворительность, автоматический анализ текстов, дискурс, библиометрическая сеть, сетевой анализ, сеть совместной встречаемости слов, социальная картография

Введение

Благотворительность издавна занимала важное место в жизни общества. Социальная значимость благотворительности заключается в том, что она обеспечивает гармонизацию интересов различных социальных и политических групп [1, 2]. Уровень развития благотворительной деятельности является одним из индикаторов социального здоровья общества, следовательно, данная тема всегда остается актуальной [3].

Благотворительность как явление постоянно трансформировалась, обретая новые смыслы и порождая новые предметно-проблемные поля в научной литературе. В разные исторические периоды благотворительность рассматривалась с различных ракурсов: с точки зрения экономики – как социальное предпринимательство [4, 5], корпоративная социальная ответственность [6, 7] или филантрокапитализм [8, 9]; с точки зрения психологии – как сфера мотивов и стимулов совершения пожертвований отдельной личностью [10–13], с позиций социологии благотворительность выступает как система отношений между людьми [14, 15].

Имеет место лавинообразный рост публикаций по благотворительности, в которых не только отражаются все новые аспекты этого социального феномена, но и происходят процессы их конструирования путем артикуляции новых значений, пропаганды новых практик и легитимизации новых форм. Этот процесс публичного «думания и говорения» о благотворительности создает огромный объем нового знания, который на данный момент представляет проблему для индивидуального усвоения. Без новых подходов к управлению знанием оно как содержательная часть всеобщего интеллекта (general intellect) человечества рискует превратиться в достояние мнемоников из известного рассказа Р. Шэкли. Перспективным направлением управления знанием выступают различные формы его визуализации, картирования и компрессии [16], позволяющие обеспечить качественную навигацию в потоках научной ин-

формации. Основой для создания этих форм являются технологии искусственного интеллекта и машинного обучения.

Цель настоящего исследования – обеспечить максимально автоматизированный анализ дискурса благотворительности, представленный в большом наборе текстов, с помощью методов библиометрического анализа с применением машинного обучения. Это позволит значительно упростить определение смыслов данного явления как в разные временные периоды, так и в разных социокультурных контекстах.

Мы предполагаем, что события последних двух лет должны нанести отпечаток, в том числе на научные публикации, формируя новые значения благотворительности. Активизируются всемирные благотворительные организации (например, Красный крест), остро встает вопрос о сборе средств для пострадавших от всемирной пандемии или для огромного количества беженцев в разные страны. Наша задача – уловить эти изменения посредством автоматического анализа большого объема научной литературы.

Научные публикации, близкие по цели к настоящему исследованию

В настоящее время литература по благотворительности, полностью соответствующая цели нашего исследования, по основным научным библиографическим базам не обнаруживается. Наиболее близкие по направленности работы (*related literature*) посвящены библиометрическому анализу литературы, связанной с узкими тематиками благотворительной деятельности, соответствующими с точки зрения однозначности анализируемого концепта требованиям, предъявляемым к составлению систематических обзоров [17, 18]. В качестве объектов (тематик) таких работ, выполненных с применением библиометрических методов, фигурируют «помощь людям, страдающим афазией» [19], «корпоративное добровольчество» [20], «корпоративная благотворительность» [21], феномен традиционного для ислама благотворительного института «*waqf*» [22]. В статье Акила Якоба и Джен Линг Ган, претендующей на анализ тематического разнообразия феноменов благотворительности на материале исследовательских работ, опубликованных в течение почти семи десятилетий [23], набор библиографических данных оказывается также ограниченным. В данном случае сужение обозреваемого поля выглядит несколько искусственным благодаря обеспечению однородности феномена (концепта) только по двум его тесно связанным лингвистическим маркерам («*charitable giving*» и «*charitable behaviour*» [23, с. 38]) в ущерб роли других десигнатов, что признают сами авторы статьи [23].

Методология исследования

В качестве обучающей выборки было решено взять научные статьи из баз данных Web of Science (WoS) и Scopus за 2020–2022 гг., содержащие в своем названии, ключевых словах и аннотациях слова, связанные с благотворительностью (*Charity, Philanthropy, Giving, Fundraising, Non-profit* и пр.). Перечисленные слова были выбраны исходя из анализа статей, опубликованных в Академии Google в период с 2020 по 2022 г. Так, например, слово *Giving* содержится в 260 000 наименованиях научной литературы, *Non-profit* – в 30 700, *Charity* встречается в названиях 33 200 статей, *Philanthropy* – в 22 000, *Fundraising* значится в названиях 21 400 научных работ. Ниже в таблице приведены слова, которые были выделены нами в качестве ключевых, а также количество научных работ, содержащих в своих названиях и аннотациях данные слова.

Для анализа дискурсивного поля благотворительности на основании загруженных статей строится тематическая карта предметной области.

Построение карты (сети совместной встречаемости слов) производится в несколько этапов. В первую очередь аннотация статьи, её название и ключевые слова объединяются в единый текст. Все полученные тексты подвергаются процессу токенизации (разбиение тек-

ста на слова, исключение из массива данных лишних знаков препинания и чисел), лемматизации (приведение слов в начальную, словарную форму), некоторые важные словосочетания объединяются в единые токены, например, red cross превращается в redcross, кроме того, к единой форме приводятся некоторые ключевые для нашего исследования слова (например, разные авторы могут писать словосочетание non-profit как nonprofit или non profit).

Количество упоминаний ключевых слов в научных работах, опубликованных в Академии Google в период с января 2020 г. по июнь 2022 г.

Ключевые слова	Количество упоминаний в Академии Google
Charity	33 200
Charitable	24 700
Philanthropy	22 000
Giving	260 000
Fundraising	21 400
Non-profit	30 700
Support	972 000
Digital charity	18 600
Charity Foundation	22 400
Donor	149 000
Donate	30 500
Donation	44 300
Crowdfunding	21 400
Altruism	21 500
Help	1 110 000
Helping	164 000
Assistance	271 000
Aid	379 000
Give	339 000
Give away	102 000
Send	124 000
Make donation	22 700
ICRC	15 000
Red cross	329 000

Следующим шагом выделяются «якорные» термины, характерные для тематик, которые точно не относятся к теме нашего исследования. Например, слово electron в словосочетании electron donation маркирует статью по физике, не имеющую отношения к благотворительности. Все тексты статей, содержащие словосочетания такого типа, считаются «мусорными» и удаляются из базы данных, чтобы не создавать смысловых искажений в сети.

Далее из базы текстов удаляются общеупотребительные стоп-слова (например, any, other, and и пр.), а также специфические слова, связанные непосредственно с написанием научных статей (такие, как research, study, result и пр.).

Из слов, прошедших первые этапы обработки, отсеиваются те, которые встречаются хотя бы в 6 разных текстах. Именно на основе этих наиболее часто употребляемых слов и строится сеть совместной встречаемости слов.

В получившейся сети могут вновь обнаружиться как новые стоп-слова, так и «мусорные» тематики, которые не удалось вычислить на первых этапах обработки слов, поэтому сеть просматривается, подвергается чистке и перестраивается. После нескольких итераций подобной чистки сеть приходит к виду, пригодному для анализа тематических полей нашей предметной области.

Результаты исследования

На рис. 1 представлена сеть совместной встречаемости слов по теме «благотворительность».

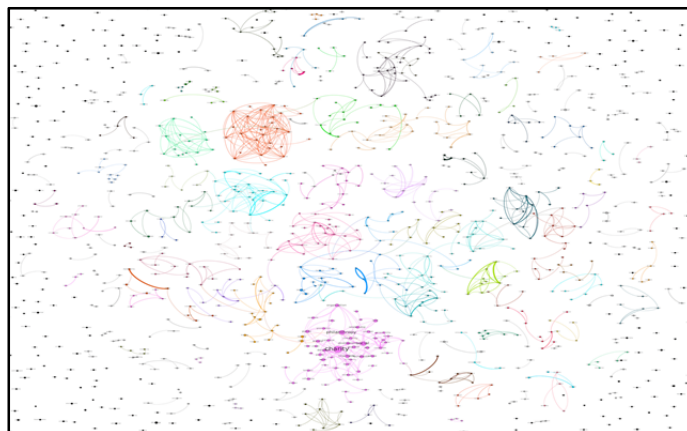


Рис. 1. Общая сеть совместной встречаемости слов

Как видно на рис. 2, центральный кластер сети составляют такие слова, как: charity, philanthropy, social, organization, charitable, foundation, community, public, people, donor, relationship и другие. Слова, образующие данный кластер, чаще всего употребляются в текстах научных статей друг с другом. Данный кластер можно охарактеризовать так: «социальная структура благотворительности». Социальную составляющую отражают такие слова, как «социальный», «отношения», «роль», «корпоративный», «общественный», «сообщество» и пр. Социальная структура благотворительности находит свое выражение в английских эквивалентах следующих слов и словосочетаний: «организация», «некоммерческий», «фонд», «фандрайзинг», «донор», «пожертвование» и пр.

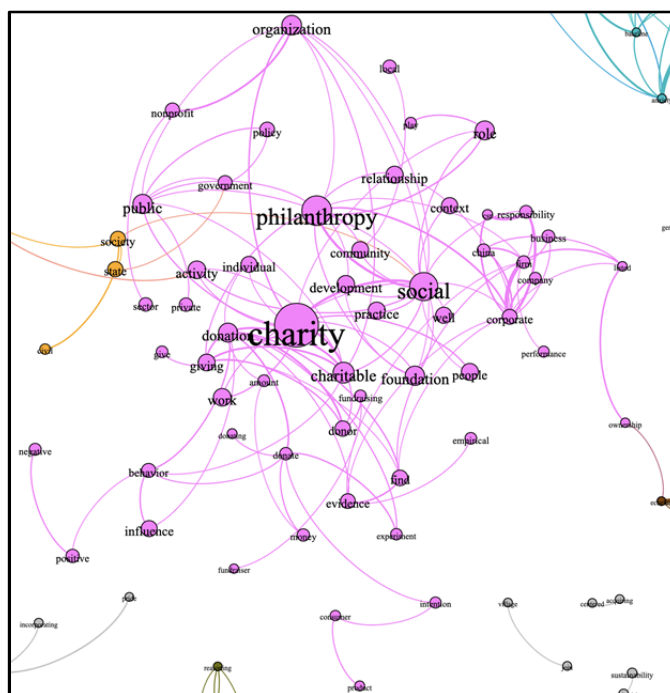


Рис. 2. Центральный кластер «Социальная структура благотворительности»

Отдельно сформировался кластер цифровой благотворительности (рис. 3), состоящий из слов: digital charity, online, collected, digital, platform, transaction, crowdfunding, campaign.

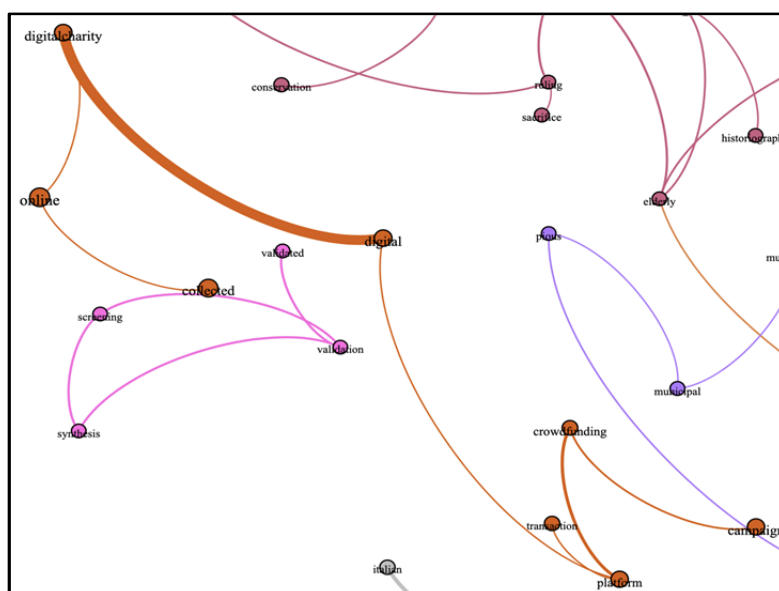


Рис. 3. Кластер «Цифровая благотворительность»

Благотворительность в научной литературе формирует обширный набор слов, связанных с медициной (рис. 4), который условно можно обозначить «Физическое и ментальное здоровье». При ближайшем рассмотрении (посредством просмотра статей, маркированных обозначенными на карте терминами) оказывается, что данный кластер объединил аспекты благотворительности разной смысловой направленности: с одной стороны, он формируется, по сути, экспансией центрального кластера «Социальная структура благотворительности» в домен медицинских учреждений (узел «hospital»), пользующихся при оказании врачебной помощи («care», «treatment», «nursing», «service») при многочисленных физических и психических заболеваниях поддержкой благотворительных фондов и/или прибегающих к прямому финансированию со стороны частных доноров для обеспечения «accessibility» и «feasibility» лечения; с другой стороны, в него оказались включенными слова (и, соответственно, статьи), относящиеся к зафиксированной еще в конце 1990-х – начале 2000-х гг. [24–26], но продолжающей в исследуемый период искать свое *особое* место в составе концепта благотворительности теме донорства «частей и продуктов человеческого тела» [27–32].

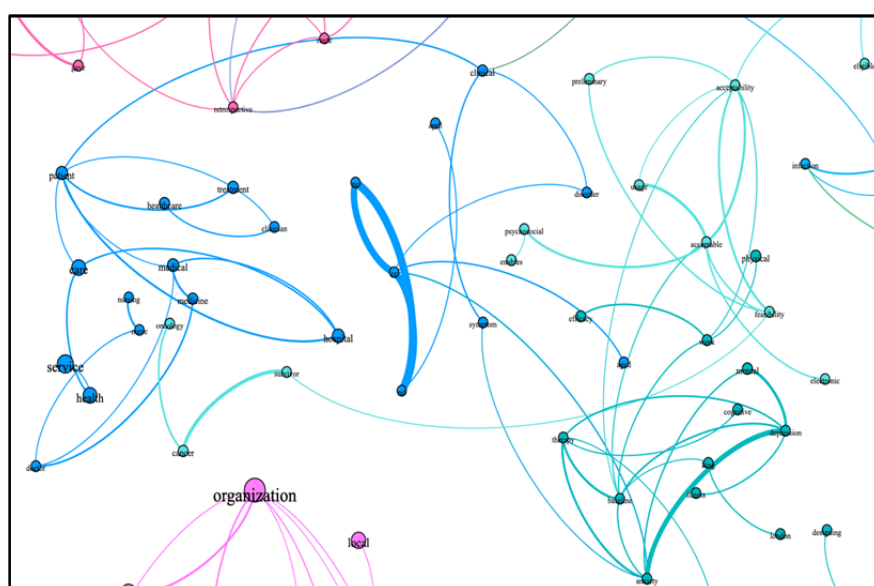


Рис. 4. Кластер «Физическое и ментальное здоровье»

Очевидно, что объединение в одном кластере исследований, относящихся, хотя и «по касательной», к «социальной структуре благотворительности», и работ, посвященных осмыслению феномена суверенности человека в отношении своего тела, представляется не вполне удачной кластеризацией, в которой формальное основание «совместной встречаемости слов» взяло верх над тематизацией на основе смысловой гомогенности.

В сформированном на основе совместной встречаемости слов поле выделяется наряду с вышеописанными явная смысловая группа слов, связанных с катастрофами и стихийными бедствиями (рис. 5): disaster, natural, hazard, flood, insurance, compensation.

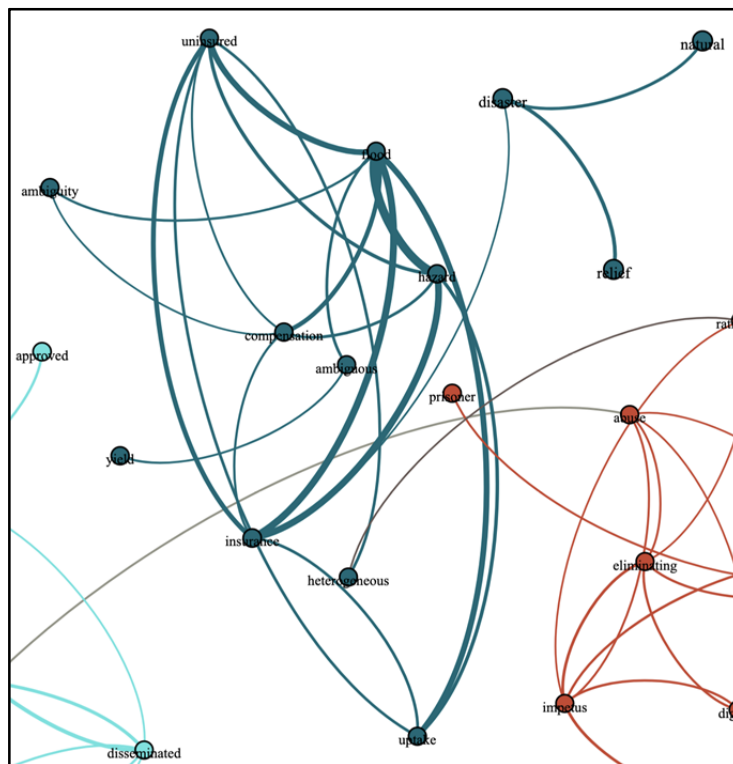


Рис. 5. Кластер «Стихийные бедствия»

Примечательно, что сформировался кластер, посвященный спонсорству спортсменов (рис. 6), визуализация которого на карте стала представлять (в смысле – «замещать») другие ранее описанные в литературе виды «благотворительности знаменитостей» (celebrity charity/philanthropy) [33–35].

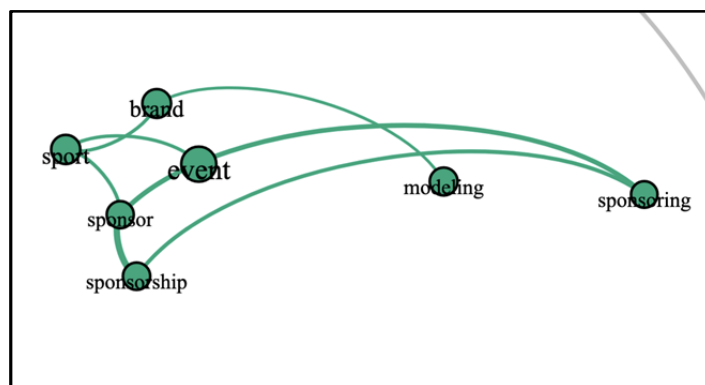


Рис. 6. Кластер «Спонсорство спортсменов»

Как видно на рис. 7, пандемия COVID-19 также нашла свое место в научной литературе, посвященной благотворительности. В данный кластер вошли такие слова, как: crisis, pandemic, covid, coronavirus, lockdown, unprecedented, disease, outbreak.

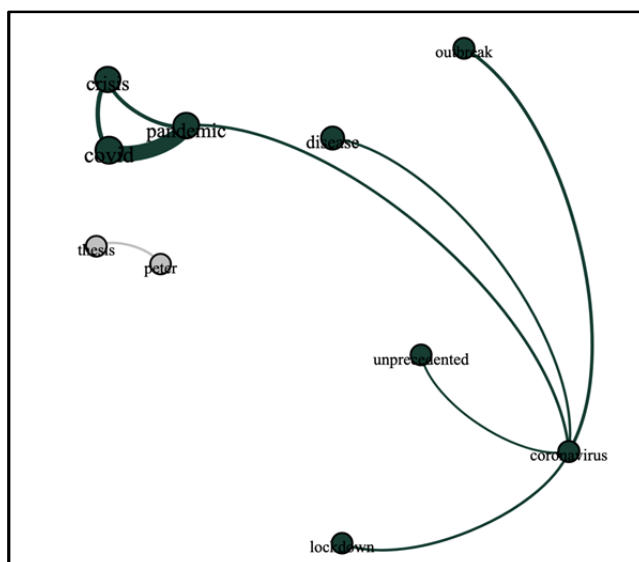


Рис. 7. Кластер «COVID-19»

Военные действия фигурируют сразу в трех кластерах (рис. 8–10): в связке с организацией Красный крест (рис. 8), в контексте проблемы беженцев (рис. 9), а также непосредственно с упоминанием таких слов, как: war, world, soldier, military и пр. (рис. 10).

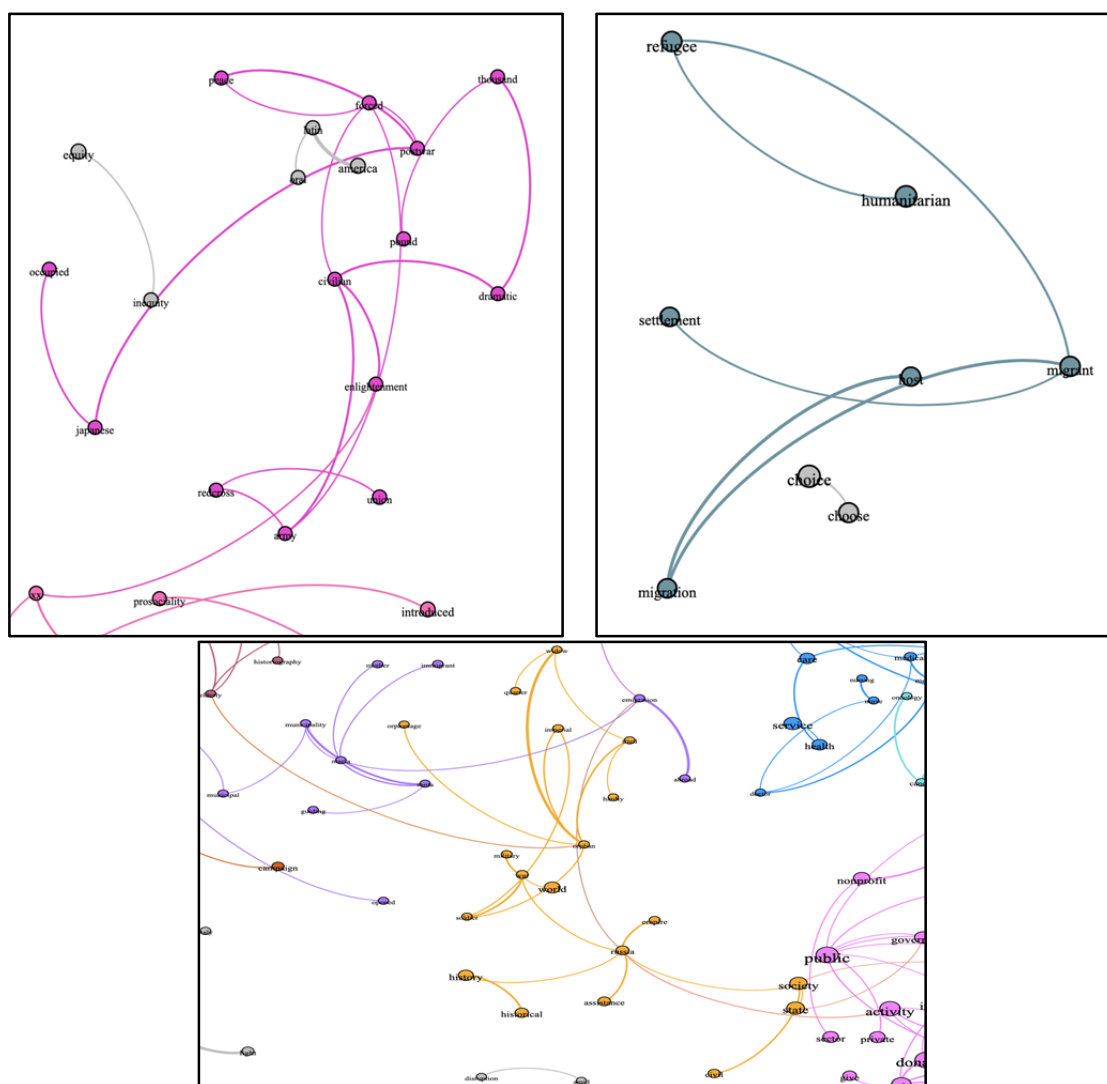


Рис. 8–10. Кластеры «Военные действия»

Заключение

Применение к текстовой базе данных, составленной из названий и аннотаций более чем 5 000 статей за 2020–2022 гг. по благотворительности, метода сетевого анализа совместной встречаемости слов позволило построить визуальную карту дифференциации дискурса. Данное построение оказалось достаточно надежным в части фиксации гегемонного дискурса (благотворительность как социальная и организационная структура общества), а также в отношении распознавания новаций, связанных с реакцией исследовательского сообщества на проявления благотворительности в условиях форс-мажорных обстоятельств (пандемия, военные конфликты) и на ее новые формы, вызванные к жизни радикальными технологическими изменениями («digital charity»).

В то же время автоматизированная дифференциация дискурса на основе анализа совместной встречаемости слов способна вводить в заблуждение, производя замену больших уже известных тем их наиболее «громкими» представителями, как это обнаружилось при рассмотрении построенного машиной кластера, условно обозначенного девизом «спонсорство спортсменов».

Случай кластера «Физическое и ментальное здоровье» демонстрирует необходимость дополнения метода анализа дискурса на основе совместной встречаемости слов инструментами более четкого анализа многозначных терминов, таких как в данном случае «медицина». Это может быть обеспечено путем «очищения» сети совместной встречаемости слов в последующих итерациях. Однако, можно предположить, что в новых итерациях рассматриваемая проблема может проявиться вновь, и место ее «проявления» достаточно сложно предугадать.

Другие способы совершенствования поиска тематически-смысловых кластеров могут быть связаны с введением интра- и экстра-дискурсивных оснований (терминология авторов) картирования дискурса, выступающих в роли координатных осей. Экстра-дискурсивный способ представлен, например, методами социальной картографии [36, 37] и методом фреймирования дискурсивного поля [38], когда *под цели* исследования (ad hoc) задаются оси, относительно которых происходит дифференциация не только тематик, но и значений социальных концептов [39].

Что же касается интра-дискурсивных координат, которые не навязывались, а *выводились* бы из дискурса, то проблема их определения находится в настоящее время только в стадии постановки [22, 40, 42].

Список источников

1. Trattner W.I. From poor law to welfare state: A history of social welfare in America. Simon and Schuster, 2007. 426 p.
2. Christianson F.Q., Thorne-Murphy L. (ed.). Philanthropic Discourse in Anglo-American Literature, 1850–1920. Indiana University Press, 2017. 255 p.
3. Ракитина А.Ю. Социальная эффективность благотворительности // Вестник Московского университета. Серия 18. Социология и политология. 2015. № 1. С. 238–244.
4. Dees J.G. A tale of two cultures: Charity, problem solving, and the future of social entrepreneurship // Journal of business ethics. 2012. Vol. 111. № 3. P. 321–334.
5. Nicholls A. ‘We do good things, don’t we?’: ‘Blended Value Accounting’ in social entrepreneurship // Accounting, organizations and society. 2009. Vol. 34, № 6-7. P. 755–769.
6. Mihaljević M., Tokić I. Ethics and philanthropy in the field of corporate social responsibility pyramid // Interdisciplinary Management Research XI. 2015. P. 799–807.
7. Leisinger K.M., Schmitt K. Corporate responsibility and corporate philanthropy // United Nations Organization. 2011. P. 1–15.
8. Bishop M., Green M. Philanthrocapitalism: How giving can save the world. Bloomsbury Publishing USA, 2010. 305 p.
9. Eikenberry A.M., Mirabella R.M. Extreme philanthropy: Philanthrocapitalism, effective altruism, and the discourse of neoliberalism // PS: Political Science & Politics. 2018. Vol. 51, № 1. P. 43–47.
10. Bekkers R., Wiepking P. A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving // Nonprofit and voluntary sector quarterly. 2011. Vol. 40, № 5. P. 924–973.

11. Mesch D.J. et al. Gender differences in charitable giving // *International Journal of Nonprofit and Voluntary Sector Marketing*. 2011. Vol. 16, № 4. P. 342–355.
12. Kosters M., Van der Heijden J. From mechanism to virtue: Evaluating Nudge theory // *Evaluation*. 2015. Vol. 21, № 3. P. 276–291.
13. McGoey L., Thiel D. Charismatic violence and the sanctification of the super-rich // *Economy and Society*. 2018. Vol. 47, № 1. P. 111–134.
14. Mersiyanova I.V., Solodova I.I. *Community Foundations in Russia: Phase of institutional development*. Publishing House of the State University – Higher School of Economics, 2010. 80 p.
15. Niumai A. Indian diaspora philanthropy: A sociological perspective // *Man In India*. 2011. Vol. 91, № 1. P. 93–114.
16. Grant M.J., Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies // *Health information & libraries journal*. 2009. Vol. 26, № 2. P. 91–108.
17. Mengist W., Soromessa T., Legese G. Method for conducting systematic literature review and meta-analysis for environmental science research // *MethodsX*. 2020. Vol. 7. P. 100777.
18. Henderson L. K. et al. How to write a Cochrane systematic review // *Nephrology*. 2010. Vol. 15, № 6. P. 617–624.
19. Asghar I., Egaji O. A., Griffiths M. An overview of the digital solutions for helping people with aphasia through bibliometric analysis // *Eneurologicalsci*. 2021. Vol. 22. P. 100311.
20. Dreesbach Bundy S., Scheck B. Corporate volunteering: A bibliometric analysis from 1990 to 2015 // *Business Ethics: A European Review*. 2017. Vol. 26, № 3. P. 240–256.
21. Mei M. Q., Wang T. Place and corporate philanthropy: A systematic review // *International Journal of Management Reviews*. 2021. Vol. 23, № 3. P. 376–393.
22. Uluyol B. et al. Mapping waqf research: a thirty-year bibliometric analysis // *Journal of Islamic Accounting and Business Research*. 2021. P. 748–767.
23. Yaacob, A., Gan, J.L. A bibliometric analysis of 68 years of research on charitable giving // *International Journal of Modern Trends in Social Sciences*. 2021. Vol. 4, № 15. P. 21–42.
24. Meslin E.M., Rooney P.M., Wolf J.G. Health-related philanthropy: Toward understanding the relationship between the donation of the body (and its parts) and traditional forms of philanthropic giving // *Nonprofit and Voluntary Sector Quarterly*. 2008. Vol. 37, № 1. P. 44S–62S.
25. Wagner L. The ‘new’ donor: creation or evolution? // *International Journal of Nonprofit and Voluntary Sector Marketing*. 2002. Vol. 7, № 4. P. 343–352.
26. Bennett R., Savani S. Factors influencing the willingness to donate body parts for transplantation // *Journal of health & social policy*. 2004. Vol. 18, № 3. P. 61–85.
27. Machin L.L., Williams R.A., Frith L. Proposing a sociology of donation: The donation of body parts and products for art, education, research, or treatment // *Sociology compass*. 2020. Vol. 14, № 10. P. 1–16.
28. Oreg A., Appe S. “Ain't No Hood Like Motherhood”: The Complexity of Human Milk Donor Identity // *Social Science Quarterly*. 2020. Vol. 101, № 2. P. 439–458.
29. Han H. J., Wibral M. Organ donation and reciprocity // *Journal of Economic Psychology*. 2020. Vol. 81. P. 102331.
30. Farsides T., Smith C. F., Sparks P. Beyond “altruism motivates body donation” // *Death Studies*. 2021. P. 1–9.
31. Oreg A., Appe S. Why is human milk donation absent from the literature on philanthropic giving? The invisible female donor and her invisible gift // *Nonprofit and Voluntary Sector Quarterly*. 2021. P. 1–12.
32. Lafuente-Funes S. Shall we stop talking about egg donation? Transference of reproductive capacity in the Spanish bioeconomy // *BioSocieties*. 2020. Vol. 15, № 2. P. 207–225.
33. Ostrower F. *Why the wealthy give: The culture of elite philanthropy*. Princeton University Press, 1997. 190 p.
34. Jeffreys E., Allatson P. *Celebrity philanthropy*. Intellect Books, 2015. 228 p.
35. Littler J. The new Victorians? Celebrity charity and the demise of the welfare state // *Celebrity studies*. 2015. Vol. 6, № 4. P. 471–485.
36. Liebman M., Paulston R. G. Social cartography: A new methodology for comparative studies // *Compare*. 1994. Vol. 24, № 3. P. 233–245.
37. Erkkila K. *Ways of Seeing the Entrepreneurial Education Debates in the United States, the United Kingdom and Finland, 1999*. 22 p.
38. Snow D. A. Elaborating the discursive contexts of framing: Discursive fields and spaces // *Studies in symbolic interaction*. 2008. P. 3–28.
39. Ferreira J.J.M., Fernandes C.I., Kraus S. Entrepreneurship research: mapping intellectual structures and research trends // *Review of Managerial Science*. 2019. Vol. 13, № 1. P. 181–205.
40. Batzdorfer V., Steinmetz H. Reviewing radicalization research using a network approach // *Journal for Deradicalization*. 2020. № 23. P. 45–95.
41. Baryshev A. The regional spectra for the entrepreneurial motives and the dissemination of entrepreneurs' knowledge in the context of modern discursive formation of entrepreneurship // *Economic and Social Development SPECIAL EDITION XIV International Conference «Russian Regions in the Focus of Changes» Book of Proceedings*. 2019. P. 154–161.

ANALYSIS OF CHARITY DISCOURSES THROUGH THE PRISM OF A BIBLIOMETRIC NETWORK

E.S. Mishchenko, A.A. Baryshev, A.V. Gubanov

National Research Tomsk State University, Tomsk, Russian Federation

Abstract. The phenomenon of charity in all epochs reflects both the level of the socioeconomic development of society and the state of moral culture of society as a whole. It also reflects civilizational maturity. This paper describes an experiment on mapping the discourse formed by recent articles which contain linguistic markers of charity: “Charity”, “Philanthropy”, “Giving”, “Fundraising”, “Non-profit”, and others. The aim of the study is to determine the possibility of identifying the thematic differentiation of charity discourse based on the construction of a network of words used in conjunction with these markers. The original database consisted of titles and annotations related to more than 5,000 articles, published in 2020–2022, found by these markers in the Web of Science (WoS) and Scopus databases. As a method of automatic text processing, the method of network analysis of word co-occurrence was used, which allowed us to obtain the following results. Firstly, a network of words and phrases was built on the basis of the Jacquard coefficient: the higher the frequency of joint use of words in relation to the total number of analyzed texts in which at least one of them occurs, the more weight these words receive, visually increasing on the map. Secondly, clusters related to various thematic areas dealing with various aspects of modern charity have been identified. These results can be useful for studying the thematic diversity of charity manifestations on a larger time scale, as well as for improving the methodology for differentiating discourses by homogeneous thematic clusters.

Keywords: charity, automatic text processing, discourse, bibliometric network, network analysis, word co-occurrence network, social cartography

ПРОБЛЕМА ИЗУЧЕНИЯ КУЛЬТУРНОГО КОДА РОССИЯНИНА С ПОМОЩЬЮ БОЛЬШИХ ДАННЫХ

Л.А. Мосунова, Е.В. Митягина

Вятский государственный университет, Киров, Россия

Исследуются элементы культурного кода содержащихся в отзывах читателей на произведения лауреатов литературных премий с 2000 по 2020 г. Цель статьи – показать процесс и предварительные результаты изучения культурного кода россиянина с помощью больших данных на основе анализа комментариев о современной литературе, опубликованных на платформе библиотеки электронных книг ЛитРес. Описаны этапы исследования: автоматический сбор данных; разработка матрицы категорий, кодификатора разметки и инструкции для аннотаторов; ручная разметка данных и создание аннотированного корпуса текстов; выявление статистически значимых закономерностей по результатам ручной разметки. Найдены подходы к контент-анализу. Результаты заключаются в анализе 7 959 текстов с точки зрения разработанной классификации. Выявлено, что 13% читателей свойственна культурная идентификация, 32% имеют социально значимую мотивацию, 49% выразили непосредственный интерес к содержанию; детально проанализирован ряд примеров. Сделан вывод, что изучение с помощью больших данных спонтанных текстов расширяет наши представления о традиционных и новых доминантах в культурном коде читателей.

Ключевые слова: культурный код, большие данные, автоматический анализ текстов, современная художественная литература, читательские комментарии

Понятие культурного кода относится скорее к метафорам, чем к научной терминологии. Отсюда бесконечное разнообразие в его определениях и подходах к изучению явления, которое за ним стоит [1, 2]. Культурный код исследуют прежде всего в социально-гуманитарных научных областях: социологии культуры, психоаналитике, семиотике, лингвокультурологии, аксиологии и др. [3–5]. Но несмотря на пестроту определений и дискуссионные аспекты его изучения, представление о культурном коде сложилось достаточно отчетливо. Учёные соглашаются, что культурный код – это то, что отличает один народ от другого, делает нас особенными; что он складывается и передаётся в традициях; что это система особых культурных координат, которые обеспечивают передачу социально значимого смысла, необходимого для взаимопонимания людей данной общности [6]. Важно, что «закодированная» система смыслов, уникальных архетипов, образов, ценностей характеризует идентичность народа – его менталитет и духовно-нравственные установки и тем самым позволяет идентифицировать культуру. Поэтому самоидентификация личности, её самоопределение неразрывно связаны с культурным кодом [7–9].

Ускорение темпов общественного развития вносит коррективы в культурный код россиян, особенно в среде молодёжи. Следовательно, проблема изучения современного состояния культурного кода россиянина остро актуальна. В научном плане *проблема* заключается в поиске методов и средств получения достоверных фактов о социально значимых смыслах жителей России на основе анализа больших данных. Нерешённой до сих пор задачей является автоматическое извлечение компонентов культурного кода из больших корпусов текстов.

Исходя из того, что к компонентам культурного кода учёные относят идентичность (этническую, национальную, региональную, религиозную), духовные ценности (нравственные, эстетические), а также знаки и символы, воплощающиеся в типичных ментальных образах [10], авторы выдвинули *гипотезу*, что составляющие культурного кода могут быть выявлены при помощи Big Data на основе анализа текстовых массивов.

По Д.А. Леонтьеву, полем коллективных смыслов является культура, а искусство – главный механизм их трансляции, обеспечивающий человеку специфический опыт [11, с. 412–438]. Непосредственная эмоциональная реакция на произведение искусства заставляет личность обнажать свой внутренний мир, его ценностно-смысловую сферу, проявлять осо-

бенности идентификации, ключевые ценности, образы, доминирующие в сознании [12, с. 115; 13, с. 330–334]. Авторы допустили, что такую реакцию можно обнаружить в письменных комментариях к современной художественной литературе, в частности, на платформе ЛитРес – крупнейшей в России библиотеки электронных и аудиокниг, занимающей 2/3 легального рынка подобной продукции и имеющей 166 тысяч читателей. Поэтому именно ЛитРес была выбрана в качестве *объекта* исследования.

Цель статьи – показать начало процесса и предварительные результаты изучения культурного кода россиянина с помощью больших данных на основе анализа массива текстов читательских откликов.

Исследование проходило в четыре *этапа*:

- 1) автоматический сбор данных;
- 2) разработка матрицы категорий, кодификатора разметки и инструкции для аннотаторов;
- 3) ручная разметка данных и создание аннотированного корпуса текстов;
- 4) выявление статистически значимых закономерностей по результатам ручной разметки;

Предмет исследования – отзывы читателей о книгах лауреатов литературных премий 2000–2020 гг.: «Большая книга», «Русский Букер», «Национальный бестселлер», «Ясная поляна», «Большая литературная премия России» и др. Отзыв понимался нами как письменное высказывание, выражающее личностное эмоционально-оценочное отношение к прочитанному. *Частная гипотеза* предполагала, что отзыв как реакция на произведение искусства содержит сведения о ценностных доминантах в сознании современного читателя, отражает актуальное состояние культурного кода, позволяет определить его изменение в условиях глобализации.

Всего в поле зрения авторов статьи попало 109 произведений премиальной литературы, принадлежащих 63 писателям.

На *первом этапе* с помощью интерфейса прикладного программирования API каталога интернет-магазина «ЛитРес» была получена информация, содержащая общие сведения: идентификатор отзыва, дату написания отзыва, имя (или псевдоним) автора отзыва, идентификатор пользователя, являющегося автором отзыва, текст отзыва, число «лайков» и «дизлайков».

Всего был получен массив из 7 959 текстов, который далее обрабатывался с помощью различных функциональных узлов программы PolyAnalyst™. Например, узел «проверка орфографии» позволил обнаружить 1 257 ошибок и при пороге достоверности 50% автоматически исправить 44% наиболее очевидных погрешностей типа *аффтар* – автор, *безисходность* – безысходность (всего 553 слова). Узел «индекс» определил исходные данные о количественном и лексическом составе каждого отзыва: количество слов, части речи, связи между словами определённой предметной области.

Важным и наиболее сложным этапом исследования стал *второй этап* – разработки матрицы категорий, кодификатора разметки и создания инструкции для аннотаторов. Исходя из теоретических основ изучения культурного кода, в итоге напряжённого интеллектуального поиска, методом проб и ошибок были выделены категории для анализа текстового массива: тип высказывания (факты, эмоции, смыслы); идентичность (я, мы, я – мы, я – они, безличный тип авторства); духовные ценности (нравственные, эстетические, нравственно-эстетические); интересы (непосредственные; опосредованные); мотивы (самовыражение, социально значимые).

На *третьем этапе* – ручной разметки данных и создания аннотированного корпуса текстов применялся метод референтного мнения – мнения эксперта, поддержанного большинством. В результате перекрёстной разметки текстов комментариев по релевантности, осуществлённой экспертом и аннотаторами, было выявлено, что 24% отзывов являются релевантными, то есть содержат элементы культурного кода. Пример использования метода референтного мнения для определения релевантности текстов приведён на рис. 1.

Эксперт	A1	A2	A3	Решение
+	-	-	-	Удаление отзыва
+	+	-	-	Удаление отзыва
+	+	+	-	Референтное мнение
+	+	+	+	Референтное мнение

Рис. 1. Разметка читательских отзывов экспертом и аннотаторами

На этом этапе произошло выделение нового, до сих пор не изученного аспекта – «оценка языка произведения в отзыве о книге как компонент культурного кода», что обозначило интересную перспективу исследования. Потребовалась также проверка новой гипотезы: частотность употребления слов-маркеров коррелирует с духовно-нравственными ценностями граждан России.

Анализ эмпирических данных по результатам ручной разметки на заключительном в настоящее время *четвёртом этапе* позволил выявить статистически значимые закономерности в проявлениях культурного кода и сделать предварительные выводы о культурных доминантах в сознании взрослой читающей публики. Интересным показалось сравнение в дальнейшем полученных результатов с ценностными ориентациями российской молодёжи в отзывах о тех же произведениях.

Так, было определено, что только 20% читателей выходят на осмысление прочитанного; люди предпочитают описывать в отзывах эмоции (51%) или перечислять факты (29%).

Приведём примеры разных типов высказываний.

Вот типично фактуальный отзыв: ««Иду на грозу» – совершенно замечательное произведение! В нём рассказывается о том, каково жить и работать в науке, с какими трудностями приходится там сталкиваться. Книга в том числе и о дерзаниях учёных, о красоте науки и окружающего нас мира. Повествование живое, герои словно невыдуманные, настоящие» (пользователь «Алёна»). В нём излагается, *о чём* книга, но не раскрывается, *что* думает автор отзыва по этому поводу.

Сугубо эмоциональная реакция воплощена в высказываниях, подобных этому: «Книга очень понравилась, невозможно оторваться, хотя сам по себе сюжет о нелёгкой судьбе татарской женщины, отправленной в ссылку, просто ужасает, тяжело читать на эту тему, стыдно за наше прошлое, за то, что так много оказалось беспринципных, безжалостных людей, готовых убивать, издеваться, предавать» (пользователь «autoreg867945370»). Доминирование эмоций подчёркивается словами «очень понравилась», «невозможно оторваться», «сюжет ужасает», «тяжело читать», «стыдно за прошлое».

Способность проникнуть в концептуальные слои художественного текста и связать образ мира писателя со своим жизненным опытом проявилась в таких примерах: «В книге нет связного захватывающего сюжета, и первые страницы я всё ждала развития сюжета. Потом поняла, что книга построена как поток воспоминаний одного человека. Воспоминаний не только личных, но и родителей и их родителей. Срез жизни нескольких поколений одной семьи в имеющихся исторических реалиях. Мне как человеку деревенскому многое показалось очень понятным и знакомым. И даже воспоминания моих родителей и бабушки во многом созвучны с написанным в книге. Читала, а сама невольно вспоминала всё, что мне рассказывали мои родственники, и удивлялась, какие у них всех были незаурядные яркие биографии» (пользователь «magueor»).

Другой важный результат был получен в отношении идентификации, являющейся ключевым компонентом культурного кода. 64% высказываний были сделаны от своего имени, 13% – в безличной форме, 10% – в противопоставлении себя другим. Представляют себя частью чего-то большего (страны, народа, нации) лишь 13%, например, как в следующем отзыве: «700 страниц на одном дыхании... столько жизней сплелись воедино... В моей семье со стороны обоих родителей были раскулаченные, враги народа. Я понимала, что это боль... но то, как это описано в этой книге... как явно описаны и прожиты все чувства, эмоции, кото-

рые пережили наши предки тогда... рекомендую однозначно к прочтению» (пользователь «Elena Zavozina»).

Интересная закономерность была выявлена при анализе мотивации и интересов читателей как характеристик наиболее глубоких проявлений культурного кода.

Социально значимые мотивы обнаружались в понимании авторами комментариев общественной значимости книги, во внутренней потребности рекомендовать её широкому кругу читателей ради общих интересов (32%). Большинство отзывов (68%) создавалось из потребности самовыражения, из желания перенести во внешний мир своё внутреннее состояние, переживания, эмоции после прочтения книги, что проявлялось в подробном описании своих эмоциональных реакций, впечатлений, ассоциаций.

Примерно в равных пропорциях проявились виды интересов читающей публики. У 49% он был непосредственным, что выражалось в единстве темы отзыва, его предметности, в конкретных деталях, выразительных подробностях, заинтересовавших читателя в книге. У 51% интерес классифицировался как опосредованный: отзыв для этих пользователей являлся не целью, а поводом для высказывания или средством для достижения какой-то иной цели, а именно, выразить своё отношение к молодому или старшему поколению, продолжить с кем-то «скрытый диалог», высказать свои взгляды на жизнь в целом.

Разработанный алгоритм классификации использования технологий машинного обучения выявил новые проблемы и позволил авторам статьи осознать *проблемную ситуацию*. Оказалось, что разработанные маркеры недостаточны для полноценного анализа текстовой информации, а предварительная оценка модели автоматического извлечения показала точность 0,765 ($\pm 0,023$) [0.743, 0.776, 0.788, 0.749, 0.770], что не соответствует эталону 0,8. Возникли новые задачи: стало необходимо уточнить подход к выбору корпуса отзывов (исследовать не просто реакцию на текст произведения, а то, каким образом это произведение прочитывается как национальное); конкретизировать компонентный состав исследуемого культурного кода, уточнить его вербальные маркеры; расширить число словарей синонимов и стоп-слов для машинной обработки текстового массива; уточнить алгоритмы поиска, апробировав их на аналогичных data-сетах [14, с. 30].

Появились *перспективы исследования* культурного кода россиянина с помощью Big Data. Это в первую очередь расширение источников данных. Ими могут стать онлайн-платформы магазинов электронных книг (MyBook, ЛитРес, Букмейт и др.), сайты книжных офлайн-магазинов (Читай-город, Лабиринт, Буквоед и др.), маркетплейсы (Ozon, Wildberries, KazanExpress и др.), социальные медиа, сообщества по интересам (VK, Одноклассники, YouTube и др.), национальный корпус русского языка (НКРЯ).

Авторы увидели также возможность расширения исследовательской базы за счёт блогов, кинофильмов, афиш, иллюстративного пропагандистского материала и других мультимодальных текстов, где исследуемый материал будет содержать, кроме вербальных (языковых) знаков, индексные, идеографические, невербальные и др. Полученный опыт позволяет использовать методы Big Data в самых разнообразных ситуациях.

Факты, полученные нами, подтверждают данные других учёных об актуальном состоянии и изменениях культурного кода россиянина в современных условиях [15, с. 25–28]. Результаты данного исследования могут быть интересны и полезны учреждениям образования и культуры, Минобрнауки, библиотекам, библиотеке машинного обучения scikitlearn, издательствам и др. Авторы планируют создать базу данных «Отзывы о художественных произведениях, содержащие сведения о культурном коде», которая отличается от аналогов процедурой формирования и тематикой, позволяющей определить типичные образы и культурные доминанты в сознании россиянина.

Список литературы

1. Бабосов Е.М. Культурный код нации: сущность и особенности // Наука и инновации. 2016. № 3 (157). С. 48–58. URL: <https://cyberleninka.ru/article/n/kulturnyy-kod-natsii-suschnost-i-osobennosti> (дата обращения: 02.06.2022).

2. Манапова В.Э. Трансформация культурных кодов: от образа – к сигналу // Общество: философия, история, культура. 2019. С. 1–5. URL: <https://cyberleninka.ru/article/n/transformatsiya-kulturnyh-kodov-ot-obraza-k-signalu/viewer> (дата обращения: 11.05.2022).
3. Аванесова Г.А., Купцова И.А. Коды культуры: понимание сущности, функциональная роль в культурной практике // В мире науки и искусства: вопросы филологии, искусствоведения и культурологии : сб. статей по материалам XLVII международной научно-практической конференции. 15 апреля 2015 г. Новосибирск : Сибирская академическая книга, 2015. № 4(47). С. 28–37. URL: <https://elibrary.ru/item.asp?id=23293205> (дата обращения: 20.05.2022).
4. Культурный код нации: Материалы семинара «Реалистическое моделирование». М. : Изд-во МГУ, 2015. 68 с. URL: http://council.gov.ru/activity/analytics/analytical_bulletins/53281/ (дата обращения: 21.04.2022).
5. Степанова Н.И. Коды культуры: семиотический и культурологический аспекты // Идеи и идеалы. 2012. № 1 (11). С. 130–135. URL: <https://cyberleninka.ru/article/n/kody-kultury-semioticheskiy-i-kulturologicheskiy-aspekty> (дата обращения: 23.05.2022).
6. Балашова К.И. Национальные культуры в контексте всемирной глобализации // Молодой учёный. 2021. № 27 (369). С. 291–294. URL: <https://moluch.ru/archive/369/83035/> (дата обращения: 10.05.2022).
7. Мосунова Л.А., Борзова Т.В. Развитие национальной культурной идентичности студентов гуманитарных специальностей в процессе обучения продуцированию нарративных текстов // Перспективы науки и образования. 2021. № 6 (54). С. 141–154. doi: <https://doi.org/10.32744/pse.2021.6.10>
8. Конева А.В., Лисенкова А.А. Матрица идентичности в цифровую эпоху: социальные вызовы преодоления анонимности // Вестник Томского государственного университета Культурология и искусствоведение. 2019. № 35. С. 14–28. URL: <https://elibrary.ru/item.asp?id=41258424> (дата обращения: 21.05.2022).
9. Гревнев В.М. Социокультурные коды как мировоззренческие и ментальные матрицы региональной идентичности // Вестник Кемеровского государственного университета культуры и искусств. 2017. № 41. С. 58–63. URL: <https://elibrary.ru/item.asp?id=32205692> (дата обращения: 22.05.2022).
10. Симбирцева Н.А. «Код культуры» как культурологическая категория // Знание. Понимание. Умение. 2016. № 1. С. 157–167. URL: <https://cyberleninka.ru/article/n/kod-kultury-kak-kulturologicheskaya-kategoriya/viewer> (дата обращения: 24.04.2022).
11. Леонтьев Д.А. Психология смысла: природа, строение и динамика смысловой реальности. М. : Смысл, 2003. 487 с.
12. Мосунова Л.А. Принцип целостного восприятия художественного текста в обучении студентов-гуманитариев // Вестник Вятского государственного университета. 2020. № 2 (136). С. 115–121. doi: <http://10.25730/VSU.7606.20.030>
13. Мосунова Л.А. Проблема воспитания эстетически развитого читателя // Берковские чтения – 2021. Книжная культура в контексте международных контактов : материалы VI Международной научной конференции (Гродно, 26–27 мая 2021 г.) / сост.: Л.А. Авгуль, Н.В. Вдовина. Минск : ЦНБ НАН Беларуси; Москва : ФГБУН НИЦ «Наука» РАН, 2021. С. 330–334.
14. Митягина Е.В., Ашихмина Т.Я., Котельников Е.В. Методы анализа социальных медиа (на примере социальной сети ВКонтакте) // Advanced Science. 2017. № 1 (5). С. 30.
15. Захарова Я.М., Зиновьева Е.Г. Культура России как фактор формирования имиджа страны // Традиционные национально-культурные и духовные ценности как фундамент инновационного развития России. 2019. № 2 (16). С. 25–27.

THE PROBLEM OF STUDYING THE CULTURAL CODE OF A RUSSIAN USING BIG DATA

L.A. Mosunova, E.V. Mityagina

Vyatka State University, Kirov, Russian Federation

Abstract. The elements of the cultural code contained in readers' reviews of the works of literary prize winners from 2000 to 2020 are investigated. The aim of the article is to show the process and preliminary results of studying the cultural code of a Russian with the help of big data based on the analysis of comments on modern literature published on the platform of the LitRes e-book library. The stages of the study are described: automatic data collection; development of a category matrix, a markup codifier, and instructions for annotators; manual data markup and creation of an annotated corpus of texts; identification of statistically significant patterns based on the results of manual marking. Approaches to content analysis have been found. The results consist in the analysis of 7959 texts from the perspective of the developed classification. The study shows that 13% of readers are characterized by cultural identification, 32% have socially significant motivation, 49% express direct interest in the content; several examples are analyzed in detail. It is concluded that the study of spontaneous texts using big data expands our understanding of traditional and new dominants in the cultural code of readers.

Keywords: cultural code, big data, automatic text analysis, modern fiction, readers' comments

МНЕНИЯ О «ЕДЕ БУДУЩЕГО» В ЗЕРКАЛЕ АНАЛИЗА ДАННЫХ СОЦИАЛЬНОЙ СЕТИ «ВКОНТАКТЕ»

М.А. Петрищева, Ю.К. Александрова

Национальный исследовательский Томский государственный университет, Томск, Россия

Изменения климата, высокое антропогенное давление на экосистему и продовольственный кризис привели, в числе прочего, к новым тенденциям в отношении пищи и питания. Одними из главных проблем, связанных с пищевой промышленностью, считаются уровень выбросов от мясной и молочной продукции, нерациональность использования земель в пищевой промышленности и темпы роста мирового голода. Для того, чтобы решить эти проблемы, исследователи и предприниматели создают Еду Будущего, которая будет одновременно сохранять полезные свойства «еды настоящего», но при этом не будет обладать ее недостатками, то есть будет экологичной, этичной, полезной, быстрой, экономичной и т.д. Однако, несмотря на то, что рынок может быть наполнен соответствующими продуктами уже сегодня, этого не происходит. Гипотеза состоит в том, что одной из причин могут являться стереотипы потребителей об инновациях в питании и недовольство качеством новых продуктов. Отражение стереотипов и проявление недовольства отражены в коммуникации в социальных медиа – они исследуются в настоящей статье на примере комментариев к постам в социальных сетях.

Ключевые слова: Еда Будущего, инновации в питании, большие данные, социальные сети, дискурс-анализ, стереотипы, мнения, паттерны мышления.

Дискурс о Еде Будущего, представленный в обсуждениях в социальных медиа, можно отнести к личностному или персональному типу общения [1]. Необходимость рассмотреть именно его заключается в том, что в знаковой системе гастрономического дискурса сконцентрированы персональная идентификация и субъективное отношение автора к предмету, то есть отношение к его вкусу [2]. Глуттонический дискурс является языковым отражением физиологических потребностей человеческого организма в области пищевых предпочтений [3]. Однако языковое выражение таких потребностей или отношения мы не сможем найти в статусно-ориентированном типе дискурса, так как он предполагает общение с точки зрения институции, к которой относятся говорящие, и в нем нет места личному отношению к предмету или личному мнению о нем [1], поэтому необходимо исследовать именно персональный тип общения.

Анализ социальных сетей наиболее показателен, так как в них представлены как краткие, оценочно-модальные сообщения, так и философские размышления авторов сообщений. Социальная сеть «ВКонтакте» выбрана нами для проведения данного этапа работы в связи с тем, что: 1) это самая популярная социальная сеть на территории Российской Федерации по данным Brand Analytics [4], следовательно, она с наибольшей точностью может отражать общественное мнение российских граждан; 2) «ВКонтакте» обладает открытым API (Application Programming Interface – «программный интерфейс приложения»), что позволяет производить выгрузку и анализ данных, находящихся в открытом доступе.

Для того, чтобы получить необходимую нам информацию, было выбраны несколько групп (пабликов), в которых публикуются в том числе сообщения (посты) о Еде Будущего, что было выявлено при анализе поля:

- 1) Фонтанка (<https://www.fontanka.ru>);
- 2) N+1 (<https://www.nplus1.ru>);
- 3) Лента.Ру (<https://www.lenta.ru>);
- 4) Мясной эксперт (<https://www.meat-expert.ru>);
- 5) РБК (<https://www.rbc.ru>);
- 6) National Geographic (<https://www.nat-geo.ru>).

Также были дополнительно выбраны три новостных СМИ, сайты которых по версии Alexa (от 27.04.2022) являются самыми популярными на территории Российской Федерации. Среди них оказались:

- 1) <https://www.rt.com>;
- 2) <https://www.gazeta.ru>;
- 3) <https://www.nsk.kp.ru>

В рейтинге Alexa также были представлены, но не выбраны нами, gia.ru и tass.ru, так как в их официальной группе отключена возможность комментирования, и [Livejournal.ru](http://livejournal.ru), так как он не является официально зарегистрированным СМИ. Также можно отметить, что выбор анализа официальных социальных сетей порталов Лента.Ру и РБК можно объяснить не только тем, что они публиковали новости по нашей теме, но и тем, что они тоже попали в рейтинг Alexa как самые популярные сайты в России.

Выгрузка данных производилась с помощью платформы для выгрузки данных социальных сетей (<https://lk.data.tsu.ru>). Данные были выгружены за последние два года, с 27.04.2020 по 27.04.2022 г. Выгрузка составила **192 521 пост**, содержащий хотя бы один комментарий от пользователей (табл. 1).

Т а б л и ц а 1

Количество выгруженных постов из пабликов «ВКонтакте»

Название СМИ	Количество постов, всего
RT	70 000
Газета.ru	51 927
РБК	45 792
КР	38 422
Лента.Ру	27 834
Фонтанка	14 191
N+1	9 054
Мясной Эксперт	4 443
NG	788

Дальнейшая работа проводилась с помощью узлов Аналитического клиента программы Megaruter PolyAnalyst 6.5. Для того, чтобы выделить и получить для анализа и последующей категоризации релевантные тексты, посвященные обсуждению инноваций в питании, были проведены следующие этапы: модификация исходных данных в машиночитаемый формат, несколько этапов чистки и фильтрации, сопровождаемые ручной перепроверкой промежуточных результатов для корректировки данных.

Отметим, что перед обработкой постов было проведено поисковое исследование, в рамках которого в новостях и научных статьях по данной тематике выделены ключевые слова и словосочетания (маркеры), относящиеся к теме «Еда Будущего». Таким образом, получено 200 маркеров для анализа языка подписчиков сообществ, среди которых с наибольшей вероятностью есть потребители Еды Будущего (люди, придерживающиеся правильного, вегетарианского, сбалансированного питания, здорового образа жизни, увлекающиеся оздоровительными практиками), и подписчики популярных новостных сообществ, которые высказывались о данной теме. Примеры маркеров: *искусственное мясо, напечатанные наггетсы, макронутриенты, лабораторное мясо, заменитель сахара, Beyond Meat, грибной белок, альтернативная еда* и другие.

В целом ключевые слова охватили следующие круги тем: заменители мясной и молочной продукции; усовершенствованные способы производства мяса и молока; внедрение обогащенных продуктов; ГМО; диеты; производители инновационных продуктов; повышение качества питания (полезные, вкусные, доступные, экологичные, быстрые, экономичные продукты).

Релевантные посты составили 0,3% от общей массы, то есть 568 постов. Операция производилась путем вычленения в текстах найденных ранее 200 ключевых слов и словосочетаний, относящихся к теме «Еда Будущего». После «ручной проверки» осталось 203 релевантных сообщения, что составило 0,1% от общей выгрузки и наглядно показывает низкий уровень обсуждаемости и интереса пользователей к данной теме в России.

Важно отметить, что после извлечения ключевых слов и «ручного отбора» в группах Мясной Эксперт и National Geographic не было выявлено ни одного релевантного поста. Это говорит о том, что несмотря на то, что новостной источник освещал проблематику Еды Будущего на своем сайте, он либо не размещал посты по данной теме в своих соцсетях, либо использовал для этого неизвестные нам лингвомаркеры, либо сообщения не были прокомментированы ни одним пользователем на странице в социальной сети. В связи с этим данные паблики исключены из дальнейшего анализа. Таким образом, релевантные посты представлены следующим образом (табл. 2).

Т а б л и ц а 2

Количество выбранных релевантных для анализа постов из пабликов «ВКонтакте»

СМИ	Количество всех постов	Количество релевантных постов	Процент релевантных постов
РТ	70 000	4	0,006
gazeta.ru	51 927	30	0,058
РБК	45 792	55	0,120
КР	38 422	61	0,159
Лента	27 834	16	0,057
Фонтанка	14 191	9	0,063
n+1	9 054	28	0,3
Мясной Эксперт	4 443	0	0
NG	788	0	0

Анализ данных показал, что больше всего постов по теме инноваций в питании выпускают новостные сообщества СМИ: Комсомольская правда (61 пост), РБК (55 постов) и Газета.ру (30 постов). Однако если говорить о том, как часто такая тема обсуждается относительно остальных информационных поводов в каждом паблике, то лидирует научно-популярное издание N+1 (0,3% от новостной повестки).

Далее были выгружены комментарии пользователей – 3 003 комментария к отобранным 203 постам, из которых после просмотра «вручную» остался 761 полезный комментарий. Из релевантных комментариев были выделены три типа сообщений:

- 1) те, в которых люди высказывают наиболее ярко выраженные в общей массе стереотипы и мнения об инновациях в питании;
- 2) те, в которых люди выражают лояльность к продукции данного типа;
- 3) остальные комментарии, содержащие сообщения о Еде Будущего, но не поддающиеся простой кластеризации.

Начнем с дискурсивного анализа первого типа комментариев с точки зрения разделения на темы и подтемы сообщений:

1. «Полезное питание – это дорого». Одним из важных критериев Еды Будущего является ее полезность. Этот критерий актуален также и для «еды настоящего», но часто не выполняется из-за высокой цены на такие продукты. Одной из целей модификации «еды настоящего» в Еду Будущего является удешевление полезных продуктов и/или изменение пищевых привычек на более здоровые и сбалансированные, которые в конечном итоге также будут более бюджетными. Однако потребители уверены, что полезные продукты в будущем не будут доступны им, как и настоящем: «...И как российскому пенсионеру прикажете сле-

дить за питанием?»; «Правильное питание – удовольствие дорогое, и не всем по зубам» (Орфография и пунктуация авторов здесь и далее сохранена).

Этот кластер комментариев является одним из наиболее радикализованных, так как с точки зрения некоторых пользователей, здоровая пища на данный момент доступна не всем слоям населения. Доступными же представляются крупы и макаронные изделия, которые, как известно, не являются полезными: «Было бы доступно по ценам здоровое питание, а так что подешевле, иногда цены на овощи и фрукты зашкаливают, поэтому курица, макароны и пастасу».

2. Второй распространенный паттерн – это мысль о том, что соя и другие растительные ингредиенты могут содержаться только в дешевых, некачественных и вредных продуктах. Также бытует миф о том, что продукты мясоперерабатывающих компаний уже сегодня являются безмясными, но это воспринимается не как инновация, а как обман населения: «Мы много лет едим растительную колбасу. Неужели другая бывает?»; «Мы уже давно перешли на растительное мясо во всяком случае в колбасных изделиях до 80% сои».

Данные примеры сообщений показывают, что соя, несмотря на все свои полезные свойства, бюджетность и перспективность для Еды Будущего, не воспринимается обществом как здоровый и натуральный ингредиент, хотя она и является растительным продуктом. Это также показывают следующие комментарии: «Опять проплаченный вброс... Брехня это всё! Соя не может полноценно заменить мясо!»; «Для начала продукты приведите в состояние НАТУРАЛЬНЫХ. А не ваши эти пальмовое масло, соя, мясо, напичканное антибиотиками, овощи со всякой х***ей и т.д. А потом заботьтесь о здоровье людей» (здесь и далее нецензурная лексика заменена символами «***»). – М.П., Ю.А.).

Существование такого стереотипа скорее всего является одной из причин медленного развития рынка растительных альтернатив мясной продукции, так как в сознании людей, видимо, существует неразделенность понятия сои как обычного бобового и сои как дешевого наполнителя.

3. Следующий паттерн, который нам удалось выделить, это боязнь ГМО. Некоторые люди уверены, что использование генномодифицированных продуктов в пищу может привести к ухудшению здоровья и снижению рождаемости: «Роман, не ври, ГМО это использование генов несвойственных генов, к примеру в картофель добавляют гены калорадского жука, медуз»; «Вы не правы и это факт Русский народ знает что ГМО это яд и вам фрицам нечего не поможет отравить русский народ!»; «Демография совсем опустится из за гмо».

Стереотипы о ГМО могут быть одним из тормозящих факторов развития рынка Еды Будущего, так как он практически неразрывно связан с генной инженерией и селекцией, однако даже те процессы и продукты, которые можно усовершенствовать без использования ГМО, воспринимаются потребителями как инновации, для которых ГМО применяется. Возможно это говорит о соединённости понятий «инноваций в питании» и «генной модификации»: «Забыли добавить ГМО» (об использовании стволовых клеток для печати куриного мяса); «ГМО-тина. А потом у людей хвост и жабры вырастут» (об использовании стволовых клеток для печати куриного мяса).

Также важно отдельно отметить, что люди уверены, что их «травят ГМО» то есть насильно заставляют употреблять продукты, содержащие ингредиенты генной модификации, а популяризаторы науки и производители такой пищи целенаправленно хотят нанести населению вред: «Успокойтесь, люди знают уже про ГМО, я считаю тех кто завозит ГМО в Россию это фашисты»; «Мрази жрите сами ГМО».

Эта тема является одной из вызывающих большое количество негатива и споров в комментариях. Однако это имеет и положительную сторону: наличие обсуждения показывает, что в сети есть достаточно большое количество пользователей, придерживающихся научной точки зрения.

4. Следующий популярный стереотип связан с тем, что люди воспринимают инновационные продукты как синтетические или химические, даже если речь идет о растительных аналогах молочной продукции, сделанных из круп и воды. «Хемофобия» – это тип страха и предубеждения по отношению к техническому процессу в области использования веществ. Обычно «химия» противопоставляется «органическим» и «натуральным» продуктам, в том числе еде, произведенной в локальных масштабах (фермерская и домашняя продукция), а не в условиях легкой пищевой промышленности [5]. То есть перед нами две проблемы:

1) Инновации в питании ассоциируются с химическим процессом, даже если им не являются: *«Молоко должно быть натуральным. А не химия растительная»*; *«Химию в массы»* (о необходимости в Еде Будущего, в связи с перенаселением планеты и, как следствие, продовольственных кризисов).

2) «Химия» в целом воспринимается как нечто негативное, вредное, уступающее по качеству «натуральному». Обычно она ассоциируется с дешевыми и вредными продуктами, которые имитируют вкусы и запахи продуктов хорошего качества: *«Сами жрите ваше синтетическое мясо»*; *«А какая теперь разница? Сейчас химическому дерьму придадут вид продуктов. Можно сразу жижу в полторашках продавать, только с разным вкусом»*.

С этим в частности связано предвзятое отношение к быстроприготавливаемой пище, которая ассоциируется в первую очередь с дешевой лапшой с имитацией вкуса мясного бульона, в то время как быстрое приготовление является также одной из черт Еды Будущего, так как экономит не только время, но и ресурсы. Химические ингредиенты воспринимаются как главный минус такой еды: *«Хаха, качественная еда быстрой не бывает. Я вот борщ варю 2.5 часа»*; *«Да и на х***н она нужна эта химия»* (о закрытии завода лапши быстрого приготовления).

При этом можно точно сказать, что скорость приготовления является ценностью для пользователей: *«Вкусняха. Для себя его открыла года два назад. Готовится очень быстро. Нежный и питательный»* (о недавно появившейся на российском рынке крупе – кускусе, которая готовится в течение одной-трех минут).

Jon Entine в своей книге о хемофобии относит этот страх к страху неизвестности, которая вызвана в частности недостаточной информированностью населения о естественнонаучном знании [5]. Одним из вариантов хемофобии является деление химических веществ на опасную и безвредную или полезную, но чаще всего такое деление ненаучно.

Существующий стереотип влияет на скорость роста рынка Еды Будущего, так как научный прогресс не может происходить без использования достижений естественных наук, однако он воспринимается как опасный.

5. Следующим паттерном является мнение пользователей о том, что цена растительных аналогов пищи определяет ее назначение: употребление растительной пищи воспринимается пользователями как дешевый некачественный тип питания, подходящий только бедным слоям населения: *«Еда будущего для бедняков»*; *«Здоровое питание»*)) *«это суррогаты для бедных»*.

В рамках этого же стереотипа можно выделить мнение о том, что растительные аналоги наоборот являются более дорогими по отношению к мясным и молочным продуктам, но в отличие от дорогих привычных продуктов они не воспринимаются как более качественные, а наоборот кажутся пользователям обманом: *«Ща зожников опять разведут на деньги»*.

Еда Будущего воспринимается людьми как недостойная замена «еды прошлого», что наглядно видно из комментариев. Стереотип сильно связан с первым, в котором люди говорят о том, что полезные продукты могут быть только дорогими.

6. Самым «политизированным» стереотипом является мнение о том, что появление Еды Будущего связано с политическими махинациями и экономическими манипуляциями привилегированных слоев населения. Пользователи подозревают наличие связи между появлением и распространением дешевых продуктов. Некоторые уверены, что дешевые аналоги будут употреблять только низшие слои населения, в то время как политические и экономические эли-

ты смогут себе позволить привычный образ жизни: *«РБК, а вы не узнавали - Путин тоже на такое мясо перешёл или как то по старинке все?...»*; *«Ну понятное дело "элитка" будет есть лобстеров, но так как лобстеров на всех не хватает, а работяги должны работать на "элитку" - будут жрать химию»*.

Еще одна группа комментариев, которую мы отнесли к данному виду стереотипа – это мнение пользователей о том, что продовольственный кризис не связан с глобальными экологическими, экономическими и гуманитарными проблемами, а связан только с политической обстановкой внутри страны или наоборот не касается страны вовсе: *«Мы же не обязаны кормить весь мир. Нам хватит наших продуктов, а остальные пусть выкручиваются как хотят:»* Лично я на заменители переходить не собираюсь, люблю натуральную еду – мясо, рыбку, овощи...»; *«Чушь, Россия способна весь мир обеспечить продуктами, если грамотны люди будут у власти»*.

Также важно заметить, что есть комментарии, в которых пользователи высказывают недоверие к финансированию таких проектов: *«Просто отрабатывают деньги, которые им выделили лоббисты определенных кругов с тугими кошельками»*.

Предположительно, такие стереотипы могли бы уменьшиться, если бы компании вели более открытую политику финансирования.

7. Седьмой кластер комментариев, который мы выделили, связан с мнением о том, что Еда Будущего не может быть вкусной и подходящей человеку. Люди, придерживающиеся такого типа питания, приравниваются к людям с психологическими и сексуальными расстройствами: *«извращения. нормальный человек эту гадость есть не будет. это отвратительно»*; *«Как можно есть эту дрянь?»*.

В этой группе также мы бы хотели отдельно выделить комментарии, связанные со стереотипом о маскулинности. Пользователи говорят о том, что уход от употребления большого количества мяса в пищу, не является нормальным для мужчины: *«Не понимаю. Иду на рынок хошь мякоть свинины - 230Р. Рёбрышки 250. Говядина мякоть 350, грудинка 260. Мужчины не будут есть траву»*; *«Безалкогольное пиво, растительное мясо, резиновые женщины. Куда катится мир?»*.

Предположительно на такой паттерн мышления можно было бы повлиять с помощью использования маркетинговых компаний, использующих популярных личностей, в том числе маскулинных мужчин, что могло бы нивелировать стереотип о том, что Еда Будущего – это причуда меньшинства.

8. Одним из самых крупных является кластер, в который мы объединили сообщения, где люди напрямую или косвенно выражают свое недоверие науке. Это могут быть ненаучные заявления, спорящие с текстом постов: *«Животный белок и витамины которые содержатся в мясе, никак нельзя заменить»*; *«Благодаря мясу эволюция происходит. Не надо ничего выдумывать»* 😊.

Также это могут быть страхи и сомнения в правоте ученых и предположения о недостаточной изученности данной сферы: *«На искусственной еде уроды вырастут»*; *«Потом как всегда окажется что это де***о вредно»*.

И сообщения, в которых пользователи выдвигают различные теории заговора: *«Просто им нужно, чтоб люди не жили более 50 лет, вот и все»*; *«Фэйк мясо от Билла Гейтса, нет уж спасибо»*.

Кроме этого отдельно можно выделить комментарии, в которых люди задаются вопросом, по какой причине ученые занимаются поиском способов создания альтернативной еды и пытаются внедрить ее в массовое употребление: *«вопрос, зачем его пробовать?»* (о растительном мясе); *«Зачем биоинженер, если мат мир создан сам случайными комбинациям без Бога- биоинженера? Пуцай и мясо само создастся»* (о создании культурального мяса).

Существование таких комментариев говорит о недостаточной степени популяризации науки.

9. Последняя группа комментариев посвящена рекламе и пропаганде. Пользователи предполагают, что сообщение, показанное им в ленте новостей оплачено, например, производителем: *«А где пометка "на правах рекламы"?»*; *«Заказная статья от производителей этого г@вна»*); *«Просто реклама очередной фигни в пищевой промышленности, проходим дальше»*.

Такой тип предположения связан с недоверием скорее не к производителю или научному сообществу, а к СМИ, в группе которого опубликован пост или к журналистике (медиа) вообще. Комментарии, в которых пользователи высказывают предположение, что информация является пропагандой, социальной рекламой или другим типом сообщения, оплаченной политическими элитами, скорее является показателем недоверия к действующей власти или оппозиции: *«Это сурковская пропаганда?»*); *«Это новость-намек для россиян, что в скором времени в России разрешат так делать, и будем мы есть мясо, сделанное из насекомых»*.

Анализ паттернов показал, что пользователи обладают большим количеством стереотипов, связанных с инновациями в питании. Это может быть одной из причин неостребованности продуктов Еды Будущего на рынке. Для того, чтобы повлиять на проблему низкого спроса на продукцию Еды Будущего, также необходимо выделить из общей массы пользователей тех, кто с наибольшей долей вероятности является или может являться потребителем инновационных продуктов. Это позволит увеличить рынок сбыта продукции нового поколения уже сегодня и расширить аудиторию потребителей. Проанализировав релевантные комментарии к постам по теме продуктов нового поколения, мы заметили, что среди них есть те, которые показывают лояльность пользователей к такой продукции. Они были кластеризованы на три группы:

Первым типом сообщений, показывающих лояльность к продукции нового поколения, являются комментарии о том, что человек или его близкие являются приверженцами определенного образа жизни или диеты, которая соответствует ценностям инновационного питания. Также в эту группу были отнесены комментарии, которые сообщали, что люди уже пробовали такую продукцию и она им понравилась: *«Если уметь готовить очень вкусно получается»*; *«Вы очень злые. Это у вас от мяса. А я веган и не нервничаю уже 20 лет»*; *«Это типа фалафелей. От мяса, естественно, отличается, но вкусно»*.

Таких комментариев меньшинство, что говорит о неперспективности рекламных компаний, направленных только на таких потребителей. Однако составление портретов таких пользователей может увеличить продажи продуктов нового поколения уже сейчас. Рекламные компании, направленные на такую группу пользователей, могли бы содержать лексику, предполагающую изначальную включенность человека в тему.

Второй тип сообщений кажется нам более перспективным, чем первый. Мы отнесли туда сообщения, в которых люди говорят о своей заинтересованности продуктом, о желании попробовать его. В том числе важно было заметить, что даже если пользователь относится скептически к продукту, он все равно готов его попробовать, если это вкусно: *«Если не страдать предрассудками, жареные кузнечики намного лучше идут под пиво, чем эти ваши чипсы и сухарики»*; *«Как будто группа для веганов... Я с удовольствием съем обычную курочку-гриль, а не это. Хотя, если вкусно, то и это съел бы конечно»*; *«Хочу таблетку Еды»*; *«Интересно попробовать!»*.

Таких комментариев больше, чем тех, которые сообщают о том, что автор является приверженцем определенной диеты или образа жизни. Составление портрета пользователя, потенциально готового попробовать продукт, могло бы оказаться перспективным, так как позволило бы расширить рынок и привлечь новых покупателей. К этому же типу можно отнести пользователей, показывающих свое одобрение продуктов нового поколения не только комментариями, но и лайками. То, что человек поставил «класс», может не означать, что он уже перешел на новый тип питания, но скорее всего означает, что человек заинтересован в теме и относится к ней положительно. Рекламные компании в таком случае могли бы быть построены на лозунгах, рассказывающих о появлении чего-то нового и/или вкусного. Лекси-

ка такого рода рекламы должна была бы исключать терминологию, которую может не знать широкий потребитель.

Последний тип лояльности, который мы выявили в комментариях под постами с темой Еды Будущего, заключался в выражении согласия с научной точкой зрения. Это были как сообщения, посвященные спору с «антинаучными» высказываниями других комментаторов, так и непосредственное вербальное выражение поддержки научному прогрессу. Также к этой группе были отнесены комментарии, в которых люди задавались вопросами или иначе показывали способность к критическому мышлению и показывали заинтересованность в данной теме с научной точки зрения: *«Интересно, кому выгодно сеять панику вокруг современных технологий, толкая страну в бездну отсталости?»*; *«2020 год а ГМО мракобесов пруд пруди»*; *«Новость хорошая, да – изначально смогут позволить себе только элиты. В статье не указано в цифрах, насколько это ускоряет и удешевит производство»*; *«Видимо, учёные вплотную приблизились к экономической эффективности процесса. Теперь найден последний штрих, и долой животноводство»*.

Пользователи, выражающие таким образом интерес к научному прогрессу в сфере пищевой промышленности, являются наиболее лояльной группой, и, очевидно, могут быть наиболее восприимчивы к информации / рекламе, рассказывающей о продуктах нового поколения с научной точки зрения. В дальнейших исследованиях на основе анализа поведения людей, оставляющих подобные комментарии, возможно составить портрет поведения человека, который придерживается научной точки зрения и обладает критическим мышлением, что позволит расширить количество потенциальных потребителей.

В целом можно заметить, что комментарии по теме Еды Будущего в основном являются негативными. Процент комментариев, которые мы определили как лояльные, невысокий – 10,2%, что, однако, выше, чем показатели в других исследованиях по этой теме, так как в последних обычно лояльной группой выбирают вегетарианцев и сторонников здорового образа жизни, которые составляют около 1% населения.

Исследование показало, что наиболее яркой проблемой личностного гастрономического дискурса о Еде Будущего сегодня является низкая информированность о достижениях научного прогресса и предубеждение насчет вкусовых качеств современных аналогов существующих продуктов питания.

На основе проведенного исследования могут быть предложены следующие варианты решения проблемы:

– Производители могут исключить из своих рекламных компаний информацию о том, что продукты не содержат ГМО и другие ненаучные мифы, а также добавить лозунги, показывающие, что Еда Будущего подходит всем, в том числе маскулинным мужчинам.

– Журналистские, предпринимательские и научные проекты о Еде Будущего могут обеспечить прозрачность финансирования и отразить это в тексте для повышения доверия к этой теме.

– При последующих исследованиях на основе проведенного нами анализа возможно составление трех типов портретов лояльных пользователей для настройки персонализированных рекламных компаний, что может увеличить рынок сбыта Еды Будущего.

Список источников

1. Карасик В.И. Языковой круг: личность, концепты, дискурс. Волгоград : Перемена, 2002. С. 116–238.
2. Федорова К.М., Руфова Е.С. Глуттонический дискурс как объект лингвистического исследования // Наука, образование и культура. 2016. № 9 (12). С. 45–48.
3. Ермакова Л.Р. Глуттонические прагматонимы и национальный характер : дис. ... канд. филол. наук. Белгород, 2011. 236 с.
4. Изменение медиапотребления русскоязычного контента на платформах соцмедиа, февраль–март 2022. URL: <https://br-analytics.ru/blog/social-media-migration-in-russia/> (дата обращения: 27.04.2022).
5. Entine J. Scared to Death – How Chemophobia Threatens Public Health // Am Cncl on Science, Health. 2011. 120 p.

OPINIONS ABOUT “FOOD OF THE FUTURE” IN THE MIRROR OF DATA ANALYSIS OF THE SOCIAL NETWORK VKONTAKTE

M.A. Petrishcheva, Yu.K. Aleksandrova

National Research Tomsk State University, Tomsk, Russian Federation

Abstract. Food is changing today and will change tomorrow; some of the reasons are climate change, high anthropogenic pressure on the ecosystem, and the food crisis. The main problems associated with the food industry scientists highlight are the level of emissions from meat and dairy products, the irrational use of land in the food industry, and the growth rate of world hunger. In order to solve these problems, researchers and entrepreneurs are creating “food of the future”, which will simultaneously meet all the criteria for “food of the present”, but at the same time will not have its negative features, that is, it will be environmentally friendly, ethical, healthy, fast, economical, and so on. However, despite the fact that the market may be filled with relevant products today, this is not happening. Our hypothesis is that one of the reasons may be consumer stereotypes about nutritional innovation and dissatisfaction with the quality of new products. Reflection of stereotypes and manifestation of discontent can be found in the language that is most pronounced in social media – comments under posts on social networks.

Keywords: food of the future, nutritional innovations, Big Data, social networks, discourse analysis, stereotypes, opinions, thought patterns

ПРОФЕССИОНАЛЬНО-ОБРАЗОВАТЕЛЬНЫЕ СТРАТЕГИИ СТУДЕНЧЕСКОЙ МОЛОДЕЖИ

И.В. Ситникова

*Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского,
Нижний Новгород, Россия*

Аннотация. Исследованы подходы к выбору профессионально-образовательной траектории студенческой молодежи. Выявлены мотивы, определяющие выбор направления подготовки и высшего учебного заведения. Установлено, что профессиональные установки и намерения реализации цели в области образования различаются в зависимости от представлений об успехе и путях его достижения. Выделены типологии профессиональных ориентаций молодежи, которые позволяют изучать особенности поведения молодых людей во время подготовки к поступлению и в период обучения в вузе. Рассмотрены перспективы применения больших данных в сфере высшего образования, в том числе возможность использования цифрового следа абитуриентов с целью повышения эффективности при наборе мотивированных первокурсников. Технологии Big Data должны стать важным инструментом при планировании и организации приемных кампаний в высших учебных заведениях.

Ключевые слова: профессионально-образовательные стратегии, профессиональный выбор, мотивация, студенческая молодежь, высшее образование, большие данные, Big Data

Источник финансирования: исследование выполнено в рамках реализации Программы стратегического академического лидерства «Приоритет 2030».

Внедрение передовых технологий Big Data происходит в самых разных отраслях, в том числе в сфере науки и образования. В российских университетах инструменты больших данных позволяют проектировать образовательные программы, прогнозировать успеваемость и посещаемость студентов, оценивать их вовлеченность в учебный процесс и научную деятельность, активность в групповой проектной работе, анализировать социальное самочувствие обучающихся, выстраивать образовательные траектории молодежи [1]. Источниками больших данных в университетах служат персональные данные, сведения о взаимодействии студентов с электронными системами обучения, показатели эффективности учебных материалов, административная и прочая информация. Использование больших массивов данных, анализ внешней и внутренней информационно-среды вуза дает возможность повысить качество образования и эффективность управления образовательными системами [2]. В вузах постоянно проводится мониторинг условий и качества образовательной деятельности, достижений обучающихся и выпускников, при этом используется сочетание различных методов сбора данных, в том числе статистические, социологические, экспертные. Социологические методы позволяют выявлять факторы, влияющие на формирование образовательной среды, определять мотивы и характер поведения студенческой молодежи, фиксировать динамику этого поведения под влиянием процессов трансформации образовательной системы, изучать спрос на образовательные услуги и образовательные предпочтения поступающих в вузы [3].

С целью изучения профессионально-образовательных стратегий молодежи в 2017–2021 гг. был проведен ряд социологических исследований, в рамках которых были проанализированы 300 глубинных интервью со студентами 1–5 курсов различных направлений подготовки 12 вузов г. Нижнего Новгорода. В результате анализа интервью были выделены три подхода к выбору профессионально-образовательной траектории молодежи: 1) профессионально мотивированные молодые люди совершают *осознанный выбор* профессии, вуза, необходимых для поступления предметов ЕГЭ; 2) многие выпускники школ совершают профессионально *неосознанный выбор*: выбирают не направление подготовки, а предметы для сдачи ЕГЭ, чаще легкие и любимые в школе, в результатах экзамена которых уверены, будущая специальность определяется данным набором госэкзаменов; 3) старшеклассники без

четкого понимания, в какой вуз и на какую специальность хотели бы поступать после окончания школы, совершают *отложенный выбор*, составляя расширенный список предметов ЕГЭ «на всякий случай». Подавляющее большинство молодежи выбирали именно вторую и третью стратегии, важным фактором для них было поступление именно на бюджетное место. Подмена выбора профессии выбором предметов ЕГЭ – весьма распространенная ошибка современных старшеклассников. Неопределившиеся в выборе профессии школьники сдают те предметы, которые хорошо знают, и на основе полученных баллов ЕГЭ ищут специальности, на которые можно поступить с таким набором экзаменов. Позднее, обучаясь в вузе на ошибочно выбранном направлении подготовки, студенты чувствуют себя не на своем месте, у них появляются трудности в учебе и желание поменять факультет. Неадекватный выбор чреват еще и тем, что такие выпускники вузов не видят своего места на рынке труда и редко работают по специальности. Сделавшие осознанный выбор студенты довольны специальностью и вузом: они поступили учиться той профессии, по которой представляют свою дальнейшую деятельность, у них есть уверенность в профессиональном будущем [4].

Данная типология мотивов профессионального выбора студенческой молодежи, построенная в результате анализа глубинных интервью, во многом соотносится с результатами анкетного опроса абитуриентов ННГУ, осуществленного на основе данных об абитуриентах, подававших документы для поступления. Количество поступающих на различные факультеты и институты ННГУ в 2018 г. составило 10 017 человек. Всем абитуриентам было предложено заполнить анкету в онлайн-режиме. В опросе приняли участие 579 абитуриентов Нижегородского государственного университета им. Н.И. Лобачевского.

Формирование образовательных стратегий и постановка профессиональных целей начинается с мотивации к образованию, которая определяется социальным окружением, влиянием ценностей, сформированных у молодых людей. Исследование мотивации абитуриентов различных факультетов/специальностей ННГУ показало, что все эти факторы многочисленны и сложно взаимосвязаны. Многомерный анализ данных анкетного опроса позволил значительно уточнить и расширить типологию качественного анализа и выделить восемь групп абитуриентов, различающихся ценностно-профессиональными установками, отношением к будущей сфере деятельности и профессии, собственными возможностями и способностями, представлениями об уровне оплаты труда [5–7].

Наиболее многочисленной является группа (60%), ориентированная на престиж профессии и образования. Все мотивы имеют области пересечения, поскольку большинство молодых людей имеют несколько мотиваций. Многие мотивы пересекаются с мотивацией престижа, который является существенным фоновым фактором и профессионального выбора, и профессионального функционирования (рис. 1).



Рис. 1. Группы абитуриентов ННГУ с различными ценностно-профессиональными установками, %

1. *Ориентация на престиж профессии и образования (60%)* – выбор авторитетного вуза с качественной подготовкой по специальности в престижной сфере деятельности.

2. *Ориентация на бизнес (27%)* – высокие заработки, карьерный рост, комфортные условия труда, возможность легко найти работу: всеми этими благами, по мнению абитуриентов, обладают бизнесмены.

3. *Желание поступить на госслужбу (26%)* связано с карьерой, но им требуются не только высокие заработки, но и уважение окружающих, гарантии социального обеспечения, хорошие условия труда.

4. *Ориентирующимся на социальное окружение – рекомендации родителей, друзей, учителей (23%)*, в общем-то, все равно, куда поступать и где учиться, лишь бы учеба была полегче, а диплом давал гарантии хорошо устроиться в жизни.

5. *Установка на личностное развитие (20%)* чаще других связана с осознанным выбором, основанным на собственных интересах и склонностях. Этим абитуриентов привлекает возможность иметь интересную и разнообразную работу, сделать карьеру, стать профессионалом.

6. *Ориентация на творчество в будущей профессиональной деятельности (8%)* – осознанный выбор определенной специальности в конкретном вузе в соответствии со своими способностями. Для них значимы единомышленники в профессиональном пространстве, они ориентированы не столько на карьеру и заработок, сколько на самореализацию.

7. *Ориентация на научно-педагогическую деятельность (12%)* – выбор, основанный на собственных интересах, что нередко коррелирует с их детскими мечтами. Здесь практически нет людей, сделавших случайный выбор, главное для них – интересная работа по уважаемой в обществе специальности, а не карьерный рост и высокая заработная плата.

8. *Молодые люди, совершившие случайный выбор (18%)*, чаще выбирали бюджетное место учебы, для поступления на которое им хватает баллов ЕГЭ. При поступлении в вуз они не знали, в какой сфере хотели бы работать, и не очень интересовались этим вопросом. Выбор большинства этой группы не является осознанным, поскольку он осуществляется часто стихийно или под влиянием окружения.

Данные типологии профессиональных ориентаций абитуриентов и студентов позволяют изучить поведение обучающихся во время подготовки к поступлению и в период обучения в вузе, а также осуществить анализ устойчивости и трансформации их ценностных установок, мотивации при выборе профессионально-образовательной траектории. Исследование цифровых следов может стать перспективным методом изучения образовательных предпочтений молодежи, дополнением к традиционному социологическому анализу. По мнению экспертов, такие исследования помогут контролировать изменение спроса на высшее образование, прогнозировать число абитуриентов в вузах, принимать управленческие решения с целью повышения эффективности деятельности учебных заведений в условиях конкурентной борьбы за талантливых студентов [8]. При этом вузы решают комплекс задач: от проектирования актуальных на рынке труда образовательных программ и прогнозирования образовательной и научной траекторий молодежи до оценки целеустремленности студентов и мотивированности профессионального выбора. Big Data в образовании позволяет анализировать информацию по каждому абитуриенту и делать выводы о поступающих на конкретную специальность в конкретном вузе и регионе, что приводит к более точному пониманию результатов приема и оптимальному планированию кампании на следующий год. Изучение в течение нескольких лет цифровых следов молодежи помогает определять тенденции спроса на образовательные программы, а университетам получать лучших для них, «своих» абитуриентов.

Список источников

1. Огурцова Е.Ю., Фадеев Р.Н. Большие данные и цифровая аналитика в университетском образовании // Ноосферные исследования. 2021. № 4. С. 37–44.

2. Гвозденко Ю.В., Ищенко А.А., Пилипенко А.В. Большие данные в системе образования // Международный студенческий научный вестник: электронный научный журнал. 2019. Вып. 5, ч. 1. URL: <http://eduherald.ru/ru/article/view?id=19731> (дата обращения: 17.05.2022).
3. Заир-Бек С.И., Мерцалова Т.А. Мониторинг в образовании как элемент системы управления // Большие данные в образовании: анализ данных как основание принятия управленческих решений : сборник научных статей I Международной конференции. 15 октября 2020 г., Москва / под общ. ред. О.А. Фиофанова. М. : Издательский дом «Дело» РАНХиГС, 2020. Гл. 1.3. С. 33–52.
4. Ситникова И.В. Специфика профессионального выбора молодежи в условиях реформы образования // Трансформация человеческого потенциала в контексте столетия / под общ. ред. З.Х. Саралиевой : в 2 т. Нижний Новгород : НИСОЦ, 2017. Т. 2. С. 588–593.
5. Иудин А.А., Ситникова И.В., Тюнтяев А.С. Влияние ценностных ориентаций на профессиональный выбор абитуриентов (часть 1) // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки. 2019. № 4. С. 89–97.
6. Иудин А.А., Ситникова И.В., Тюнтяев А.С. Влияние ценностных ориентаций на профессиональный выбор абитуриентов (часть 2) // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки. 2020. № 1. С. 42–51.
7. Иудин А.А., Ситникова И.В., Тюнтяев А.С. Влияние ценностных ориентаций на профессиональный выбор абитуриентов (часть 3) // Вестник Нижегородского университета им. Н.И. Лобачевского. Серия: Социальные науки. 2020. № 4. С. 61–70.
8. Габдрахманов Н.К., Орлова В.В., Александрова Ю.К. Цифровой след в прогнозировании образовательной стратегии выпускников школ // Университетское управление: практика и анализ. 2021. Т. 25, № 3. С. 6–13.

PROFESSIONAL AND EDUCATIONAL STRATEGIES OF STUDENT YOUTH

I.V. Sitnikova

*Lobachevsky State University of Nizhny Novgorod – National Research University,
Nizhny Novgorod, Russian Federation*

Abstract. The author analyzes approaches to the choice of professional and educational trajectory of students and reveals motives that determine the choice of training direction and institution of higher education. The author learns that professional attitudes and intentions for realizing educational aims differ depending on the perception of success and ways of its achievement; typologizes young people’s professional orientations that allow investigating the peculiarities of young people’s behavior during their preparation to the university and during the period of studying in the university; and considers the prospects of applying Big Data in higher education, including the possibility of using applicants’ digital footprints to improve efficiency in recruiting motivated first-year students. Big Data technologies should become an important tool in planning and organizing admission campaigns in higher education institutions.

Keywords: vocational and educational strategies, professional choice, motivation, students, higher education, Big Data

ИДЕНТИФИКАЦИЯ ЯЗЫКА ПОЛЬЗОВАТЕЛЕЙ «ВКОНТАКТЕ»: РАЗВЕДОЧНЫЙ АНАЛИЗ

Д.В. Стефановский¹, П.А. Басина²

¹ *Российский государственный гуманитарный университет, Москва, Россия*

² *Национальный исследовательский Томский государственный университет, Томск, Россия*

Аннотация. Согласно данным статистики, 43% населения мира знают и свободно владеют двумя языками; это явление называется билингвизм. Однако билингвизм представляет скорее социокультурный феномен – человек, владеющий несколькими языками, также владеет и основами культур языковых общностей. При этом важно понимать, что люди используют разные языки не только в разговорной речи, но и при взаимодействии в социальных сетях в повседневной жизни. Цель данной работы заключается в том, чтобы провести разведочное исследование по определению языка пользователей социальной сети «ВКонтакте». В дальнейшем это поможет разработать методологию для того, чтобы идентифицировать эмигрировавших пользователей. Выбор социальной сети, в свою очередь, обусловлен несколькими факторами: популярность среди русскоязычной аудитории, публичный API, временная детализация контента. В качестве данных выступили сообщения, опубликованные на открытых стенах пользователей за 2019–2020 гг., которые в качестве страны проживания указали одну из следующих: Литва, Армения, Чехия. Программное обеспечение для сбора данных реализовано на скриптовом языке программирования Python. Выбор стран обусловлен различиями по географической принадлежности, национальному составу. Для идентификации языка использованы библиотеки LangDetect и CLD2.

Ключевые слова: идентификация языка, машинное обучение, социальная сеть «ВКонтакте», большие данные, LangDetect, CLD2

На сегодняшний день билингвизм является масштабным языковым явлением. Это значит, что многие современные люди владеют двумя различными языками. Например, согласно статистическим данным, 43% населения мира являются билингвами, а 17% – мультилингвами, то есть говорят более чем на двух языках [1, 2]. Например, в Европейском Союзе 23 официально признанных языка; 98% жителей Люксембурга говорят на родном языке и еще одном дополнительном, а в Словении и Литве этот показатель достигает 92% [3].

Исследователи выделяют несколько причин масштабов данного явления. Во-первых, важно понимать, что количество языков в мире значительно превышает количество стран. Например, в 2011 г. на 193 страны приходилось 7 000 языков [4]. Во-вторых, влияют процессы иммиграции и эмиграции. Неоспоримым фактором, влияющим на количество языков, которыми владеет человек, является образовательный процесс и воспитание. Например, студенты могут проходить обучение в другой стране, осваивая при этом язык принимающей страны [5]. Или ребенок воспитывается в семье, где родители являются носителями разных языковых культур [6].

Однако билингвизм представляет скорее социокультурный феномен. То есть человек, который владеет несколькими языками, также владеет и основами культур языковых общностей [7]. Люди используют разные языки не только в разговорной речи, но и при взаимодействии в социальных сетях в повседневной жизни. При этом, для того, чтобы эффективно анализировать эти данные, необходимо идентифицировать язык [8]. Определение языка пользователей открывает следующие возможности:

1. Идентификация родного языка [9, 10].

Основная задача данного направления состоит в том, чтобы идентифицировать родной язык пользователей на основе их публикаций на втором (изучаемом или приобретённом) языке. Основная гипотеза заключается в том, что характеристики родного языка будут проявляться в речи пользователей на другом языке.

2. Определение языка для выявления различных социальных явлений – оскорбительные высказывания в социальных сетях, проявления агрессии и т.д. [11, 12].

3. Идентификация языка является первичным шагом для эффективного применения методов обработки естественного языка.

Цель данной работы заключается в том, чтобы провести разведочное исследование по определению языка пользователей социальной сети «ВКонтакте». В дальнейшем это поможет разработать методологию для того, чтобы идентифицировать эмигрировавших пользователей. Выбор социальной сети обусловлен следующими факторами: популярность среди русскоязычной аудитории, публичный API, временная детализация контента [13, 14].

В качестве данных выступили сообщения, опубликованные на открытых стенах пользователей за 2019–2020 гг., которые в качестве страны проживания указали одну из следующих: Литва, Армения, Чехия. Программное обеспечение для сбора данных реализовано на скриптовом языке программирования Python; оно имеет ряд модулей, в частности, для работы с API «ВКонтакте», записи результатов в хранилище и обеспечения параллелизма при выгрузке. Для хранения выгрузок используется СУБД PostgreSQL. Для того чтобы получить необходимые данные сначала были найдены идентификаторы всех пользователей, указавших Армению, Литву или Чехию в качестве страны проживания. По полученным идентификаторам были выгружены открытые данные стен этих пользователей. Найдены они были уже в готовой выгрузке пользователей «ВКонтакте»; процесс сбора данных происходил в период с осени 2019 г. по август 2020 г.

Выбор стран обусловлен их различиями по географической принадлежности, национальному составу. Армения расположена на границе Европы и Азии. Согласно данным переписи за 2021 г., численность населения составляет 2 972 008 человек; количество пользователей из Армении, опубликовавших пост на своей странице «ВКонтакте», за исследуемый период равно 122 107. Отличительная особенность данной страны заключается в ее мононациональности: «97% населения представлено исконными армянами или иными этносами, причисляющими себя к этой народности» [15]. Согласно статистическим данным, в стране проживает 11 000 русских, что представляет 0,4% общего населения государства [15].

Чехия находится в центре Европы. Согласно данным статистики за 2019 г., численность населения составляет 10 707 242 человека. Количество пользователей, опубликовавших пост на своей странице «ВКонтакте», составляет 221 956. Исследователи отмечают, что значимым источником прироста населения стали иммиграционные процессы, где в качестве основной причины выступает получение образовательных услуг. Согласно данным 2021 г., на территории проживает 45 000 русских [16, 17, 18, 19, 20].

Литва расположена в северной Европе. Согласно данным переписи населения за 2021 г., в стране проживает 2 669 443 человека. Например, годом ранее данный показатель составлял 2 740 859. Количество пользователей, опубликовавших пост на своей странице «ВКонтакте», 88 171. Исследователи отмечают демографический кризис в стране. Литва является многонациональной страной. Согласно переписи населения за 2011 год, на территории проживает 176 913 русских [21, 22].

Общая задача определения языка состоит в том, чтобы предсказать для текста язык, на котором он написан. Типичный наивный подход к решению данной задачи заключается в том, чтобы показать текст определенному эксперту по языку, который затем определит, на каком языке он написан. Однако в таком случае для решения задачи потребовалось бы привлечение множества различных экспертов по каждому из языков. Это решение становится еще более проблематичным, если база текстовых данных обширна и не статична, а сами тексты написаны на нескольких разных языках. Все это создает предпосылки использовать машинное обучение, что позволяет сократить временные и трудовые ресурсы для решения задачи.

Задача предварительной обработки данных обычно является первым этапом в процессе анализа данных с помощью машинного обучения. На первом шаге происходит удаление всех ссылок и выражений – обращений к конкретному пользователю (форма @username удаляется из каждого текстового сообщения с применением регулярных выражений, поскольку они считаются нерелевантными для различения языков. Кроме того, удаляются все эмодзи,

так как они имеют одинаковую форму на разных языках. Затем текст преобразуется в строчные буквы и удаляются все знаки препинания. Эта процедура преобразовывает все тексты в одинаковый формат.

На сегодняшний день существуют различные библиотеки на скриптовом языке программирования Python, которые позволяют определить язык, на котором написаны тексты. Каждый из подходов отличается с точки зрения функционала и точности при определении того или иного языка. В рамках данной работы были использованы две библиотеки – LangDetect [23] и CLD2 [24]; они представляют собой две предварительно обученные системы. То есть для их использования нам не нужно заранее обучать модель, показывая правильные ответы о том, к какому языку относится то или иное сообщение. Например, библиотека LangDetect поддерживает 55 языков; а CLD2 – 83 языка. LangDetect представляет собой наивный классификатор Байеса, использующий представление на основе n-грамм символов без выбора признаков, с набором эвристики нормализации для повышения точности. CLD2 – идентификатор языка, встроенный в веб-браузер Chrome – использует классификатор наивного Байеса и стратегии токенизации, специфичные для скриптов. В результате тестирования библиотек лучшие результаты были получены с использованием CLD2. Важно отметить, что результаты, полученные CLD2, значительно улучшаются при применении методов предварительной обработки, указанных выше, по сравнению с результатами на сырых данных.

Рассмотрим полученные результаты.

Для Чехии популярными языками с точки зрения общего количества символов являются: русский (9 169 161), чешский (1 730 389), английский (695 197), украинский (212 306) и сербский (67 877). При этом популярными языками с позиции среднего количества символов – русский (208), финский (191), белорусский (147), литовский (142) и украинский (135). Самыми распространёнными среди пользователей языками являются – русский, чешский и английский; популярные языковые сочетания – русский и чешский, русский, чешский и английский, русский и английский.

Для Армении популярными языками с точки зрения общего количества символов выступают – русский (7 978 759), армянский (831 101), английский (378 564), сербский (41 659), чешский (16 169); по среднему количеству символов – йоруба (312), казахский (273), русский (170), кхаси (158), эстонский (117). В ТОП-3 самых частотных языков среди пользователей входят – русский, армянский и английский; распространённые языковые сочетания – русский и армянский; русский, армянский и английский; русский и английский.

Для Литвы популярными языками с позиции общего количества символов являются – русский (10 931 521), английский (391 835), литовский (231 606), сербский (34 998) и белорусский (14 828). С точки зрения среднего количества символов – русский (237), казахский (167), белорусский (156), украинский (132) и македонский (126). В ТОП-3 самых распространённых языков среди пользователей входят – русский, английский и литовский; языковые сочетания – русский и английский; русский и литовский; русский, литовский и английский.

На рис. 1 представлено распределение количества языков, используемых пользователями в зависимости от страны проживания. В случае Чехии большинство пользователей знают 3–4 языка; похожая картина наблюдается и для Армении. В случае Литвы – пользователи используют чаще 2 или 3 языка.

Подведем итоги. Социальная сеть «ВКонтакте» – полезный источник для поиска соотечественников, проживающих за рубежом. Выявленные с помощью алгоритмов машинного обучения популярные языковые сочетания – русский, английский и официальный язык страны проживания. Полученная информация позволяет понять языки, с помощью которых можно вести диалог с соотечественниками.

При этом возникает несколько вопросов при использовании данного подхода. Во-первых, возникает необходимость определения «рабочих» библиотек – то есть эффективных решений для автоматического определения языка. Это требует дополнительных тестирований. Во-вторых – необходимость определения устойчивых выражений для того, чтобы

исключить их из дальнейшего анализа; например, «тама тіа». В-третьих, важно разработать эффективную методологию валидации полученных результатов для определения того, насколько точно модели идентифицирует язык текстов.

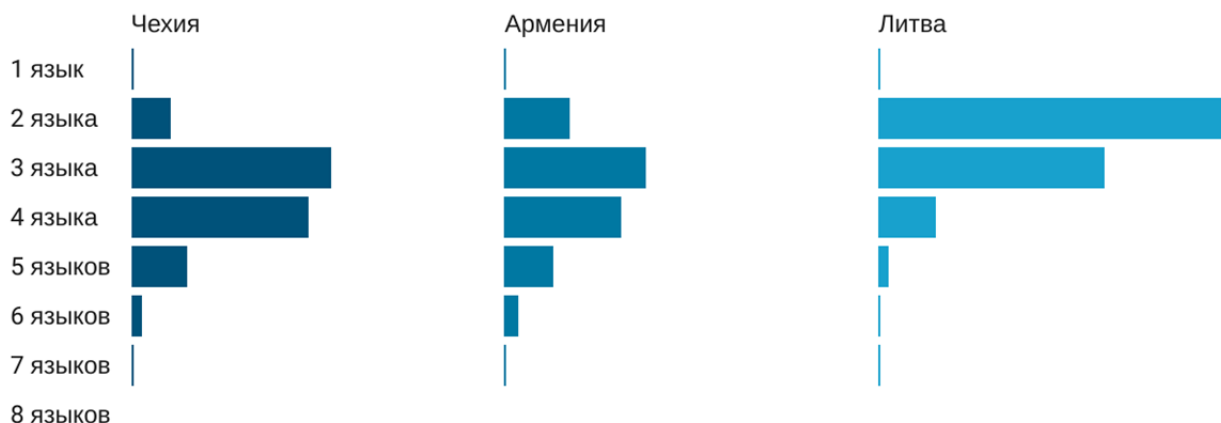


Рис. 1. Распределение количества используемых языков пользователями в зависимости от страны проживания

В дальнейшем мы планируем продолжить данную работу с учетом выделенных вопросов и протестировать методы, используемые в рамках направления определения родного языка пользователей. Также мы предполагаем, что будет эффективна стратегия анализа данных с фокусированием внимания на том, чтобы интерпретировать полученные результаты с учетом характеристик пользователей.

Список источников

1. Gration E. Bilingualism in 2021: US, UK & Global Statistics // Preply. URL: <https://preply.com/en/blog/bilingualismstatistics/#:~:text=There%20are%20approximately%2033%20billion,to%20learn%20in%20the%20US> (дата обращения: 01.04.2022).
2. Multilingual People // Ianguages. URL: <https://ilanguages.org/bilingual.php> (дата обращения: 02.04.2022).
3. Nardelli A. Most Europeans can speak multiple languages. UK and Ireland not so much // The Guardian. URL: <https://www.theguardian.com/news/datablog/2014/sep/26/europeans-multiple-languages-uk-ireland> (дата обращения: 01.04.2022).
4. Gordon R. Ethnologue: languages of the world. Dallas, Tex : SIL International, 2005.
5. Кошель Т.В. Проблема формирования и функционирования билингвизма в семье и обществе // Publishing house «Sreda». URL: <https://phsreda.com/e-articles/79/Action79-53624.pdf> (дата обращения: 01.03.2022).
6. Серова Т.С. Сбалансированный билингвизм и механизм языкового переключения в устной переводческой деятельности в условиях диалога языков и культур // Язык и культура. 2010. № 4 (12). С. 44–56.
7. Абдулина Г.Р. К вопросу о формах и результатах взаимовлияния языков // Вестник Челябинского государственного университета. 2008. № 3. С. 5–9.
8. Sarma N., Singh S.R., Goswami D. Word Level Language Identification in Assamese-Bengali-Hindi-English Code-Mixed Social Media Text // International Conference on Asian Language Processing (IALP). 2018. P. 261–266.
9. Markov I., Nastase V., Strapparava C. Exploiting native language interference for native language identification // Natural Language Engineering. 2020. Vol. 28. P. 167–197.
10. Markov I., Nastase V., Strapparava C. Anglicized Words and Misspelled Cognates in Native Language Identification // BEA@ACL. 2019. P. 275–284.
11. Косми Т., Bojar O. LanideNN: Multilingual Language Identification on Character Window // ArXiv. 2017. Vol. abs/1701.03338. P. 927–936.
12. Frey J., Stemle E.W., Doğruöz, A.S. Comparison of Automatic vs. Manual Language Identification in Multilingual Social Media Texts // Building computer-mediated communication corpora for sociolinguistic analysis. 2019. P. 47–69.
13. Digital 2022: THE RUSSIAN FEDERATION. // DATAREPORTAL. URL: <https://datareportal.com/reports/digital-2022-russian-federation> (дата обращения: 01.03.2022).
14. Щекотин Е.В., Коварж Г.Ю., Гойко В.Л., Петров Е.Ю., Бакулин В.В. Оценка качества жизни населения регионов РФ на основе цифровых данных: методологические аспекты // Векторы благополучия: экономика и социум. 2020. № 3 (38). С. 138–156.

15. Население Армении: численность, состав, динамика по годам // Статистика и показатели: региональные и федеральные. URL: <https://rosinfostat.ru/naselenie-armenii-i-statisticheskaya-informatsiya/#i-4> (дата обращения: 02.04.2022).
16. Data - počet cizinců // ČSÚ. URL: https://www.czso.cz/csu/cizinci/4-ciz_pocet_cizincu#cr (дата обращения: 02.04.2022).
17. Ethnic composition of Lithuania 2011 // Population statistics of Eastern Europe & former USSR. URL: <http://popstat.mashke.org/lithuania-ethnic2011.htm> (дата обращения: 02.04.2022).
18. Население Чешской Республики // Статистика и показатели: региональные и федеральные. URL: <https://rosinfostat.ru/naselenie-cheshskoj-respubliki/> (дата обращения: 05.04.2022).
19. Чехия // Кноема. URL: <https://кноема.ru/atlas/Чехия> (дата обращения: 02.04.2022).
20. Официальное название страны: Чешская Республика // Чувашская республика: официальный портал органов власти. URL: <http://gov.cap.ru/home/15/Chez/All.htm> (дата обращения: 02.04.2022).
21. Численность населения Литвы – 2020 // Sputnik. URL: <https://lt.sputniknews.ru/20200311/Chislennost-naseleniya-Litvy--2020-11558350.html> (дата обращения: 02.04.2022).
22. Население Литвы // Migrantumir. URL: <https://migrantumir.com/naselenie-litvy> (дата обращения: 02.04.2022).
23. Langdetect // PyPi. URL: <https://pypi.org/project/langdetect/> (дата обращения: 02.04.2022).
24. Cld2 // PyPi. URL: <https://pypi.org/project/cld2-cffi/> (дата обращения: 02.04.2022).

IDENTIFICATION OF THE LANGUAGE OF VKONTAKTE USERS: AN EXPLORATORY ANALYSIS

D.V. Stefanovsky¹, P.A. Basina²

¹ *Russian State University for the Humanities, Moscow, Russian Federation*

² *National Research Tomsk State University, Tomsk, Russian Federation*

Abstract. According to statistics, 43% of the world's population knows and is fluent in two languages. This phenomenon is called bilingualism. However, bilingualism is rather a sociocultural phenomenon. Bilinguals know the basics of cultures of linguistic communities. People use different languages not only in colloquial speech, but also when interacting on social networks in everyday life. The aim of this work is to conduct an exploratory study to determine the language of users of the VKontakte social network. In the future, this will help to develop a methodology for identifying expatriate users. The choice of a social network is due to several factors: popularity among the Russian-speaking audience, public API, temporary content detail. The data were messages published in 2019–2020 on the open walls of users who indicated one of the following as their country of residence – Lithuania, Armenia, Czech Republic. The data collection software is implemented in the Python scripting programming language. The choice of countries is due to differences in geographical affiliation, national composition. The LangDetect and CLD2 libraries were used to identify the language.

Keywords: language identification, machine learning, VKontakte, big data, LangDetect, CLD2

ИССЛЕДОВАНИЯ НА «БОЛЬШИХ ДАННЫХ»: ПРИМЕНЯТЬ ЛИ АГРЕГАЦИЮ В ЦЕЛЯХ МОДЕЛИРОВАНИЯ?

В.М. Тимирьянова, И.А. Лакман, Г.Р. Баймурзина

Башкирский государственный университет, Уфа, Россия

Аннотация. Целью исследования было определение влияния степени агрегации данных на возможности выделения пространственных эффектов в рамках прогнозных моделей. Исследование проводилось на высокочастотных обезличенных в соответствии со статьей 4.1 Федерального закона № 54-ФЗ данных о покупках хлебобулочных изделий, предоставленных АО «Энергетические системы и коммуникации» (Первый ОФД), детализированных в минутах и часах. В материалах представлены результаты проведенных тестов на стационарность данных, наличие ненаблюдаемых эффектов, последовательную корреляцию и кросс-секционную зависимость, наличие случайных и пространственных эффектов и др. Приведены оценки расчета модели с фиксированными эффектами для двух вариантов детализации данных. Результаты расчетов показали, что модель, построенная на данных, детализированных по часам, лучше, чем модель, построенная на данных, детализированных по минутам. Обе модели сходятся в своих оценках пространственных эффектов, что говорит о том, что пространственные зависимости на рынке существуют и могут быть учтены при принятии решений.

Ключевые слова: высокочастотные данные, обезличенные фискальные данные, пространственные эффекты, пространственные модели на панельных данных, большие данные

Источник финансирования: исследование выполнено в рамках государственного задания Министерства науки и высшего образования Российской Федерации (код научной темы FZWU-2020-0027).

Научно-технологический прогресс определил широкие возможности для сбора, обработки и анализа «больших данных» в целях принятия решений. Многочисленные исследования показывают многообразие возможностей использования больших данных в целях прогнозирования различных показателей, характеризующих развитие на микроуровне (например, домохозяйств, предприятий) и макроуровне (например, городов, регионов, стран) [1, 2]. Одновременно исследователи задают контрольные вопросы: в какой степени большие данные на самом деле приводят к улучшению прогнозных моделей? Когда большие данные действительно лучше? [3] В частности, приводятся доводы о том, что с данными большой размерности связаны вычислительные и статистические проблемы, включая масштабируемость, накопление шума, ложную корреляцию, случайную эндогенность и ошибки измерения [4]. Декларируется, что «детализированные данные часто оказываются очень изменчивыми из-за воздействия различных случайных факторов, разброса значений и, поэтому слабо отражают общие тенденции и свойства исследуемого множества» [5]. При этом приводятся примеры как увеличения эффективности прогнозирования при повышении детализации данных, так и обратных ситуаций [3]. На наш взгляд, такая дискуссионность возникает в силу различия применяемых инструментов, задач исследования и, если углубиться в проблему, то неоднозначности внутренней логики существующих зависимостей в данных. Последнее фактически определяет то, насколько вообще нужно для принятия решений учитывать в моделях зависимости, проявляющиеся на определенном уровне детализации.

Для исследования этого вопроса сосредоточились на конкретном виде данных и типе анализа. В качестве информационной базы рассматривались высокочастотные обезличенные данные о покупках, предоставленные АО «Энергетические системы и коммуникации» (далее Первый ОФД) по соглашению о научно-техническом сотрудничестве в обезличенной форме в соответствии со статьей 4.1 Федерального закона № 54-ФЗ. Рассматриваемые данные, характеризующие только одну группу товаров (хлебобулочные изделия), представлены в максимальной детализации по шкале времени: с точностью до секунды. По географическому

признаку данные представлены в разрезе зон г. Уфы, обеспечивающих конфиденциальность данных и невозможность выделения отдельных предприятий. Таким образом, данные позволяют установить, в какой зоне г. Уфы и какой хлеб приобретается населением, с точностью до секунды. На основе таких данных можно прогнозировать изменение объема продаж во времени и пространстве, для чего могут быть использованы инструменты панельного и пространственного регрессионного анализа.

Безусловно, задачи даже в рамках прогнозирования продаж могут быть совершенно разными. Применительно к рассматриваемому набору данных, это может быть анализ продаж конкретного вида, конкретной торговой марки хлеба или хлебобулочных изделий в целом. Интерес может представлять конкретная торговая зона, в том числе с выделением ее типов: спальный район, административный, культурный и т.д. Значительное внимание может быть уделено анализу продаж в динамике, в том числе их цикличности (рис. 1).

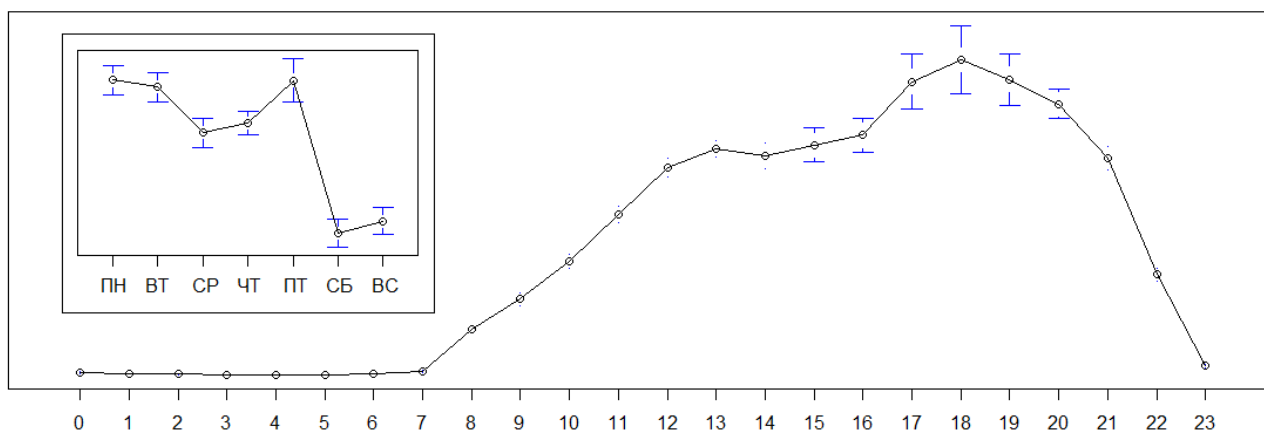


Рис. 1. Внутрисуточная и внутринедельная сезонность в продажах хлебобулочных изделий в г. Уфе.

Источник: рассчитано авторами по обезличенным данным, предоставленным Первым ОФД

Однако если опуститься на уровень конкретной марки хлеба, то есть максимально детализироваться по видам товаров и максимально детализироваться по шкале времени, то встает вопрос, насколько значимыми являются данные о секунде, в которую был реализован товар. Одновременно возникает вопрос, как такая детализация повлияет на возможности выделения пространственных эффектов, поскольку сложно себе представить, что какая-либо марка хлеба будет ежесекундно реализовываться в соседних зонах. Соответственно, в условиях анализа разреженной матрицы данных, для которой характерны преимущественно нулевые значения элементов, поиск зависимостей будет сильно усложнен.

Чтобы понять, как на результат анализа может повлиять степень детализации данных, были проведены различные статистические тесты. Расчеты проводились на стандартизованных данных о продажах хлеба в г. Уфе. Представленные в табл. 1 тесты на стационарность панельных данных показали различные результаты для данных, сгруппированных по часам и минутам. Стационарность означает, что статистические свойства процесса, генерирующего ряд, не изменяются со временем. Все проведенные тесты для данных, сгруппированных по часам, показали, что нулевая гипотеза об отсутствии стационарности отвергается в пользу альтернативной. В то же время тесты для данных, сгруппированных по минутам, указывали на наличие ложной регрессии по времени.

Далее были проведены статистические критерии, позволяющие определить наличие ненаблюдаемых эффектов в уравнениях панельных регрессий, последовательную корреляцию, кросс-секционную зависимость остатков и т.д. Результаты данных тестов с указанием альтернативных гипотез представлены в табл. 2.

Для данных, сгруппированных по часам, все тесты, кроме одностороннего теста Бера-Соса-Эскудеро-Юна, указывают на наличие последовательной корреляции данных при $p < 0,05$, а следовательно, на существование связи значений показателя с их значениями в предыдущие пе-

риоды времени. В то же время аналогичные тесты, проведенные на данных, детализированных по минутам дополнительно, показали на отсутствие последовательной корреляции в исходных ошибках. Одновременно для обоих наборов данных тесты показали наличие кросс-секционной зависимости, указывающей на наличие связи между наблюдениями. Одной из причин возникновения такой зависимости может быть наличие пространственно-связных эффектов в данных. Результаты тестирования на наличие пространственных эффектов представлены в табл. 3.

Т а б л и ц а 1

Результаты тестирования гипотезы о нестационарности данных

Наименование теста	Детализация	Простая модель	Индивидуальный перехват	Индивидуальный перехват и временной тренд
Тест Левин-Личу на наличие единичных корней	Часы	$z = -28,6;$ $p\text{-value} < 0,0001$	$z = -66,7;$ $p\text{-value} < 0,0001$	$z = -148,6;$ $p\text{-value} < 0,0001$
	Минуты	$z = -543,79;$ $p\text{-value} < 0,0001$	$z = -1097,9;$ $p\text{-value} < 0,0001$	$z = -1814,2;$ $p\text{-value} < 0,0001$
Модифицированный тест Чои на единичные корни	Часы	$P_m = 108,8;$ $p\text{-value} < 0,0001$	$P_m = 254,5;$ $p\text{-value} < 0,0001$	$P_m = 306,7;$ $p\text{-value} < 0,0001$
	Минуты	$P_m = 14,5;$ $p\text{-value} < 0,0001$	$P_m = -7,48;$ $p\text{-value} = 1$	$P_m = -7,48;$ $p\text{-value} = 1$
Тест Чои на обратные нормальные единичные корни	Часы	$z = -35,9;$ $p\text{-value} < 0,0001$	$z = -57,6;$ $p\text{-value} < 0,0001$	$z = -65,6;$ $p\text{-value} < 0,0001$
	Минуты	$z = \text{Inf};$ $p\text{-value} = 1$	$z = \text{Inf};$ $p\text{-value} = 1$	$z = \text{Inf};$ $p\text{-value} = 1$

Примечание. Выбор длины лага осуществлялся на основе модифицированного критерия Шварца.

Источник: рассчитано авторами по обезличенным данным, предоставленным Первым ОФД.

Т а б л и ц а 2

Результаты тестов на наличие ненаблюдаемых эффектов, последовательную корреляцию и кросс-секционную зависимость

Наименование теста	Значение		Вывод
	Данные, детализированные по часам	Данные, детализированные по минутам	
Тест Бера–Соса–Эскудеро–Юна на AR(1)-серийную корреляцию	$\chi^2 = 9959,7; df = 1;$ $p\text{-value} < 0,0001$	$\chi^2 = 20140; df = 1;$ $p\text{-value} < 0,0001$	Наличие авторегрессии первого порядка (AR(1)) в остатках
Односторонний тест Бера–Соса–Эскудеро–Юна на ненаблюдаемые эффекты	$z = -12,7;$ $p\text{-value} = 1$	$z = -6,7;$ $p\text{-value} = 1$	Отсутствие авторегрессии первого порядка (AR(1)), случайные эффекты не значимы
Совместный тест Балгати и Ли на случайные эффекты и последовательную корреляцию при нормальности и гомоскедастичности идиосинкразических ошибок	$\chi^2 = 9987,8; df = 2;$ $p\text{-value} < 0,0001$	$\chi^2 = 20168; df = 2;$ $p\text{-value} < 0,0001$	Наличие последовательной корреляции и/ или случайных эффектов
Тест Вулдриджа на последовательную корреляцию в моделях с фиксированными эффектами	$F = 504,6; df_1 = 1;$ $df_2 = 20102;$ $p\text{-value} < 0,0001$	$F = 134,2; df_1 = 1;$ $df_2 = 1209542;$ $p\text{-value} < 0,0001$	Наблюдается последовательная корреляция
Тест Вулдриджа на последовательную корреляцию в ошибках первой разности	$F = 73,1; df_1 = 1;$ $df_2 = 20046;$ $p\text{-value} < 0,0001$	$F = 170206; df_1 = 1;$ $df_2 = 1209486;$ $p\text{-value} < 0,0001$	Наблюдается последовательная корреляция в ошибках первой разности
Тест Вулдриджа на последовательную корреляцию исходных ошибок	$F = 50,45; df_1 = 1;$ $df_2 = 20046;$ $p\text{-value} < 0,0001$	$F = 0,061622; df_1 = 1;$ $df_2 = 1209486;$ $p\text{-value} = 0,804$	Наблюдается последовательная корреляция в исходных ошибках только в данных, детализированных по часам
Тест Песарана на кросс-секционную зависимость	$z = 465,57;$ $p\text{-value} < 0,0001$	$z = 499,28;$ $p\text{-value} < 0,0001$	Наблюдается кросс-секционная зависимость
LM Тест Бройша Пегана на кросс-секционную зависимость	$\chi^2 = 268483;$ $df = 1540;$ $p\text{-value} < 0,0001$	$\chi^2 = 381270;$ $df = 1540;$ $p\text{-value} < 0,0001$	Наблюдается кросс-секционная зависимость

Источник: рассчитано авторами по обезличенным данным, предоставленным Первым ОФД.

Тесты на наличие случайных и пространственных эффектов

Наименование теста	Значение		Вывод
	Данные, детализированные по часам	Данные, детализированные по часам	
Тест Балтаги–Сонга–Коха H0: $\sigma_{\mu}^2 = \rho = 0$ против альтернативы, что $\sigma_{\mu}^2 \neq 0$ или $\rho \neq 0$	$LM-H = 13551$; $p\text{-value} < 0,0001$	$LM-H = 14305$; $p\text{-value} < 0,0001$	Присутствуют случайные региональные эффекты и/или пространственная автокорреляция
Тест Балтаги–Сонга–Коха на маржинальный эффект H0: $\sigma_{\mu}^2 = 0$ при предположении, что $\rho = 0$ против односторонней альтернативы, что $\sigma_{\mu}^2 \neq 0$	$LM_1 = -5,3$; $p\text{-value} = 2$	$LM_1 = -5,3$; $p\text{-value} = 2$	Отсутствуют случайные эффекты
Тест Балтаги–Сонга–Коха на маржинальный эффект H0: $\rho = 0$ при условии отсутствия случайных эффектов ($\sigma_{\mu}^2 = 0$), против альтернативы что $\rho \neq 0$	$LM_2 = 116,4$; $p\text{-value} < 0,0001$	$LM_2 = 119,6$; $p\text{-value} < 0,0001$	Присутствует пространственная автокорреляция

Источник: рассчитано авторами по обезличенным данным, предоставленным Первым ОФД.

Проведенные тесты указывают на наличие пространственной зависимости данных, что обуславливает целесообразность использования инструментов пространственной эконометрики.

Применяя различные методы оценки пространственной автокорреляции ошибок, было построено несколько моделей, где в качестве зависимой переменной выступало стандартизованное в пределах географических зон значение продаж хлебобулочных изделий, а в качестве объясняющей переменной день недели (0 – выходной день, 1 – будний день). Акцент сделан на пространственно-панельных моделях, в рамках которых возможно выделение пространственных эффектов. В табл. 4 представлены результаты расчета модели с фиксированными эффектами.

Таблица 4

Результаты построения пространственной модели с фиксированными эффектами

Наименование показателя	Данные, детализированные по часам	Данные, детализированные по минутам
День недели, β_1	0,0442362 * (0,0091966)	0,0291507* (0,0014107)
Пространственный лаг, ρ	0,0021679* (0,00001495)	0,0014* (0,0000065)
Пространственная ошибка, λ	-0,00088383* (0,00004972)	-0,0012668 * (0,0000101)
Псевдокоэффициент детерминации R^2	0,44	0,008
Логарифм правдоподобия	-22676,1	-1711271
Критерий Айкаике (AIC)	45360,2	3422550
Критерий Шварца (BIC)	45391,9	3422598

* $p < 0,00001$.

Источник: рассчитано авторами по обезличенным данным, предоставленным Первым ОФД.

В первую очередь следует отметить невысокую предсказательную способность моделей. На текущий момент в открытом доступе находится ограниченное число данных в аналогичной детализации, что оказывает влияние на спектр учитываемых в моделях факторов. Сравнение моделей по псевдокоэффициенту детерминации R^2 показывает, что модель, построенная на данных, детализированных по часам, лучше, чем модель, построенная на дан-

ных детализированных по минутам. Таким образом, подтвердилось предположение о том, что чрезмерная детализация может негативно сказываться на возможностях выделения зависимостей. В то же время, согласно данным, представленным в таблице, обе модели сходятся в своих оценках пространственных эффектов. Значения коэффициентов статистически значимы ($p < 0,00001$), что говорит о том, что пространственные зависимости на рынке существуют. Положительное значение коэффициента пространственного лага ρ указывает на наличие прямой связи. Таким образом, рост продаж наблюдается на фоне роста продаж в соседних территориях. С другой стороны, коэффициент пространственной ошибки λ отрицательный. Следовательно, существует фактор, не включенный в модель, рост которого негативно сказывается на изменении объема продаж на соседних территориях. Будний день оказывает положительное влияние на рост продаж. Невысокие значения коэффициентов пространственного лага и пространственной ошибки указывают на то, что они не являются определяющими, по крайней мере, для данного вида продукта в данном промежутке времени. Низкое значение псевдо R^2 определяет слабую предсказательную возможность модели. Очевидно, что для повышения предсказательной мощности моделей требуется включение дополнительных факторов, что актуализирует поиск источников новых данных в той же детализации, для включения в модели. В то же время для целей принятия решения можно учесть выделенные зависимости, связанные с направлением (прямым или обратным) влияния включенных в модель факторов.

Представленный сравнительный анализ результатов позволяет говорить о том, что более высокая детализация данных, а именно переход на минуты, негативно сказалась на прогностических способностях выстраиваемых моделей. Об этом также сигнализировали результаты проводимых предварительных статистических тестов. Таким образом, применительно к пространственно-панельным моделям повышение детализации данных не обязательно положительно скажется на качестве прогноза. Данный вывод фактически согласуется с мнением ученых Сибирского федерального университета о том, что проблема прогнозирования «существенно усложняется, если необходимо исследовать большие пространственно-временные ряды», для которых «в силу их большого объема достаточно сложно выявить существующие в них закономерности, оценить изменчивость и построить адекватную модель для решения задачи прогнозирования» [5]. В то же время более высокая агрегация (например, по дням или месяцам, как в данных Федеральной службы статистики), безусловно, приводит к потере значимой информации о внутридневных колебаниях. Одновременно, по результатам текущего исследования следует отметить использование значительных вычислительных мощностей при построении пространственно-панельных моделей, значительно превышающих потребности для построения просто пространственных или просто панельных моделей. В связи с этим нельзя не согласиться с мнением о том, что «работа с большими наборами пространственных данных требует разработки и реализации новых статистических инструментов» [5]. Остается актуальной потребность в разработке алгоритмов выбора наиболее подходящей степени детализации и инструментов агрегации при работе с большими данными. Очевидно, что в ряде случаев необходимо формирование алгоритмов отсека наборов данных по географическому критерию, так как пространственные эффекты затухают с увеличением расстояния между наблюдениями.

Список источников

1. Гурская Е.Д., Доценко М.А., Соколянский В.В. Технологии Big Data в сервисе: новые рынки, возможности и проблемы // Вопросы экономических наук. 2015. № 4 (74). С. 42–44.
2. Митрович С. Рынок «больших данных» и их инструментов: тенденции и перспективы в России // МИР (Модернизация. Инновации. Развитие). 2018. Т. 9, № 1. С. 74–85.
3. Junqué de Fortuny E., Martens D., Provost F. Predictive Modeling With Big Data: Is Bigger Really Better? // Big Data. 2013. № 1 (4). P. 215–226.
4. Fan J., Han F., Liu H. Challenges of Big Data analysis // National Science Review. 2014. Vol. 1, № 2. P. 293–314.

5. Добронев Б.С., Попова О.А., Вюст К.В. Подходы к исследованию больших временных рядов спутниковых данных // Региональные проблемы дистанционного зондирования Земли : материалы VIII Международной научной конференции; электронное научное издание / науч. ред. Е.А. Ваганов, отв. ред. Г.М. Цибульский. Красноярск, 2021. С. 69–72.

BIG DATA RESEARCH: SHOULD AGGREGATION BE USED FOR MODELING?

V.M. Timiryanova, I.A. Lakman, G.R. Baymurzina

Bashkir State University, Ufa, Russian Federation

Abstract. The aim of the study was to determine the impact of data aggregation on the ability to identify spatial effects within predictive models. The study was conducted on high-frequency anonymized (in accordance with Federal Law No. 54-FZ) data on purchases of bakery products provided by Energy Systems and Communications (First OFD), detailed in minutes and hours. The materials present the results of tests for data stationarity, unobservable effects, sequential correlation and cross-sectional dependence, random and spatial effects, etc. Fixed effects models for two variants of data aggregation are presented. The calculation results showed that the model built on data detailed by hours is better than the model built on data detailed by minutes. Both models converge in their estimates of spatial effects, which suggests that spatial dependencies exist in the market and can be taken into account when making decisions.

Keywords: high-frequency data, depersonalized fiscal data, spatial effects, panel data spatial models, big data

Financial support: This research was funded by the Ministry of Science and Higher Education of the Russian Federation (Scientific Code FZWU-2020-0027).

БОЛЬШИЕ ДАННЫЕ В КРОСС-КУЛЬТУРНЫХ ИССЛЕДОВАНИЯХ: ДОВЕРИЕ К ИНФОРМАЦИИ И ДОВЕРИЕ К РЕЗУЛЬТАТАМ

М.С. Фабрикант

Белорусский государственный университет, Минск, Беларусь

В статье рассматриваются возможности и ограничения использования больших данных в кросс-культурных исследованиях. Прослеживается, каким образом применение больших данных позволяет решить ряд методологических и процедурных проблем, характерных для кросс-культурных исследований, и какие новые трудности при этом возникают. Делается общий вывод о том, что выбор между специально сгенерированными и большими данными отражает различные стороны проблемы доверия к исследованию: в первом случае на первый план выходит проблема доверия к самим данным, во втором – к результатам исследования. Проблема доверия к результатам будет возрастать по мере демифологизации алгоритмов машинного обучения, что, в сочетании с распространением информации об угрозах кибербезопасности, может привести в том числе к обострению межкультурных противоречий при организации кросс-культурного исследования.

Ключевые слова: большие данные, кросс-культурные исследования, цифровой след, социальное доверие, доверие к информации

Финансовая поддержка: исследование выполнено в Нижегородском государственном университете имени Н.И. Лобачевского в рамках реализации программы стратегического академического лидерства «Приоритет-2030».

Введение

Переход к использованию больших данных – методологическая трансформация, которая, как ожидается, затронет все области исследований в социальных науках, радикально изменив способы сбора данных и критерии оценки их качества, а также будет способствовать дальнейшему усложнению методов статистического анализа данных. Вместе с тем пока эти изменения продолжаются, не ясно, когда они завершатся, а что более важно, чем именно – ожидаемым методологическим переворотом, с которым связываются надежды на повышение репрезентативности и валидности выводов и улучшение предсказательной силы моделей, или не в первый раз уже переоцененным по масштабу и последствиям попыткам решить внутренние методологические проблемы социальных наук путем следования заданным извне трендам. Вероятнее всего, большие данные, опять-таки как это уже было с прежними методологическими инновациями, продемонстрируют лишь избирательную эффективность и, соответственно, позволят решить лишь некоторые проблемы и окажутся в одних предметных областях исключительно необходимыми, в других – лишь немногим более эффективнее привычных подходов, а в третьих – лишь временной данью методологической моде. Поэтому имеет смысл рассматривать перспективы использования больших данных в социальных науках дифференцированно – исходя из особенностей различных областей, причем не только очерченных дисциплинарными границами.

Цель данной статьи – проанализировать возможности и ограничения использования больших данных в кросс-культурных исследованиях. Кросс-культурные исследования интересны тем, что именно кросс-культурный поворот во многом задал предыдущие изменения, сопоставимые по ожидавшемуся от них масштабу с теми надеждами, которые сейчас возлагаются на переход к большим данным, и также в связи с проблемой воспроизводимости результатов в различных условиях. Первоначально кросс-культурные исследования фигурировали преимущественно как отдельные, достаточно специфические направления внутри отдельных дисциплин – кросс-культурная психология, сравнительная социология, сравнительная политология, – привлекавшие тех исследователей, кого интересовали межкультурные и межстрановые различия. Однако впоследствии было обнаружено, что многие феномены, ко-

которые до того считались универсальными, общечеловеческими, и изучались на данных, собранных в одной стране (как правило, к тому же, на достаточно специфической по базовым социально-демографическим параметрам подвыборке населения этой страны) с последующей экстраполяцией на человечество в целом, на самом деле во многих культурах не воспроизводятся и универсальными не являются. Поэтому универсальность каких бы то ни было предполагаемых закономерностей постепенно перестала восприниматься как опция по умолчанию [1]. Доказательство этой предполагаемой универсальности, как и выявление кросс-культурных различий, требует проведения кросс-культурных исследований не только там, где ожидается обнаружить существенные различия между культурами, но и там, где исследователи рассчитывают, что таких различий выявлено не будет. Поэтому современные кросс-культурные исследования – это и трансдисциплинарная область, для которой вопрос организации сбора репрезентативных данных стоит особенно остро, и методологическая перспектива, появившаяся непосредственно перед поворотом к большим данным и сопоставимая с ним по масштабу.

В данной статье будут рассмотрены вначале возможности, которые использование больших данных открывает в кросс-культурных исследованиях, затем – ограничения, связанные с различными форматами их использования, и в заключение – предположительные пути преодоления этих ограничений. Основной тезис заключается в том, что обращение к большим данным в кросс-культурных исследованиях позволяет существенно повысить доверие к данным, которое, однако, не обязательно конвертируется в доверие к самому исследованию и его результатам.

Возможности использования больших данных в кросс-культурных исследованиях

Одним из основных преимуществ больших данных является то, что они по определению не требуют специально организованной процедуры для их сбора. Это преимущество особенно значимо для кросс-культурных исследований, поскольку очевидно, что любое кросс-культурное исследование предполагает сбор данных в более чем одной стране – от нескольких до десятков стран. Парадоксальным образом, если для исследования, в фокусе которого находятся культурные особенности, может быть необходимо и достаточно нескольких стран, каждая из которых представляет особый интерес для исследователя, то для исследования универсалистского типа, замысел которого предполагает отсутствие межкультурных различий, требует выборки, которая на страновом уровне может обоснованно претендовать на репрезентативность. Иными словами, именно те исследователи, которые не заинтересованы в культурно-специфических сторонах изучаемого явления (зато заинтересованы в их отсутствии), зачастую вынуждены прилагать больше организационных усилий для генерирования данных, собранных у представителей различных культур.

Эти организационные усилия не сводятся к большему объему работы, что создает проблемы для исследователей, однако само по себе не сказывается на качестве данных. Кросс-культурный формат исследования означает новые, дополнительные препятствия на пути к обеспечению репрезентативности данных и их эквивалентности. Для того чтобы исследование было репрезентативным на культурном уровне, оно должно охватывать достаточно большое количество стран, которые могут существенно различаться по условиям для проведения сбора данных, особенно посредством массовых опросов общественного мнения (уровень грамотности населения, готовность участвовать в опросах и содержательный характер мотивации, особенности законодательства, регулирующие проведение опросов и других исследований с участием людей и хранение полученных данных), а также по количеству и уровню подготовки исследователей, способных и готовых участвовать в сборе данных [2].

Обычно эти проблемы решаются одним из двух способов – в зависимости от объема финансирования. Кросс-культурные исследования с объемом финансирования, достаточным для проведения полноценного, то есть на репрезентативной выборке, опроса или исследова-

ния с использованием другого метода сбора данных (серии экспериментов, интервью, фокус групп и др.: вопреки все еще существующим представлениям, использование качественных методов не исключает требования репрезентативности, хотя и предполагает иные критерии ее оценки [3]) опираются на достаточные ресурсы, чтобы обеспечить большой объем данных внутри каждой единицы анализа, соответствующей культурному уровню анализа (обычно речь идет о странах, несмотря на обвинения такого подхода в методологическом национализме). Однако качество этих данных может существенно различаться. Основным организатор исследования вынужден либо доверять организациям, проводящим сбор данных в каждой стране, либо каким-то образом контролировать качество уже собранных данных. Второе затруднительно по той же причине, по которой трудно оценить качество данных по результату безотносительно процедуре их сбора: какими бы контринтуитивными ни оказались результаты, не исключено, что проблема не в качестве данных, а в интуиции исследователя и в предсказательной силе теоретических моделей, на которые она опирается. Первый вариант – доверие страновым организациям-партнерам – требует либо надежной и детальной инсайдерской информации о десятках стран, которой главный организатор исследования обычно не располагает, либо адекватной процедуры отбора. При организации этой процедуры исследователи часто сталкиваются с тем, что более информативные критерии оценки качества (например, понимание неочевидных нюансов методологии, важных для процедуры исследования, или честность в предоставлении данных) оказываются наиболее сложными в применении, поскольку требуют все той же инсайдерской информации, а наиболее простые критерии (например, общая стоимость сбора данных, представленная в смете) – наименее информативными.

Кросс-культурные исследования с небольшим или вовсе отсутствующим финансированием – это, как правило, академические проекты, которые предпринимаются отдельными учеными, без опоры на институциональный ресурс. Преимущество таких исследований – простота контроля за качеством данных в той мере, в которой оно зависит от уровня компетентности тех, кто проводит сбор данных в каждой стране. Обычно речь идет о задействовании личных академических контактов в сочетании с методом «снежного кома» при отсутствии институциональных ограничений, которые со стороны сложнее оценить, чем компетентность отдельного исследователя. Однако при этом в силу ограниченных ресурсов сбор данных проводится не только на выборке меньшего объема, но и, как правило, среди тех, кто оказывается наиболее доступен для участия в исследовании – круг знаковых исследователя либо студенты, если исследователь преподает (так называемая *convenience sample*). Помимо внутристрановой нерепрезентативности, такой подход также ставит под вопрос кросс-культурную эквивалентность данных: социальный статус, материальное положение и другие важные характеристики студентов в разных странах могут существенно различаться. Кроме того, из-за организационных трудностей внешнего характера и недостаточных ресурсов для их оперативного преодоления сбор данных в разных странах может проводиться в разное время, вследствие чего становится невозможным дифференцировать культурные особенности и влияние событий, происходивших в период сбора данных.

Использование больших данных позволяет решить одни из этих проблем и полностью избежать других. Так, отсутствие специальной процедуры сбора данных снимает необходимость заручиться личным согласием каждого из участников исследования, поскольку речь идет о данных, которые генерируются людьми в ходе повседневной активности. Поэтому ни репрезентативность выборки, ни качество процедуры генерирования данных не страдают от описанных выше причин, связанных с особенностями самих исследователей. Кроме того, использование больших данных делает полностью нерелевантной еще одну проблему, специфическую для кросс-культурных исследований, – кросс-культурную валидизацию инструментария, что, разумеется, далеко не сводится к задаче перевода вербальной составляющей инструментария на языки задействованных стран [4]. Наконец, при наличии в команде исследователей, владеющих достаточным количеством языков, можно обойтись без страновых

партнеров на местах, привлечение которых для решения организационных вопросов иначе было бы совершенно необходимо.

Таким образом, на первый взгляд, создается впечатление, что большие данные – если не панацея, то, по крайней мере, исключительно эффективный способ сделать данные, используемые в кросс-культурных исследованиях, одновременно более качественными и более доступными для ученых с различным уровнем финансирования, и единственное препятствие на пути к реализации этой возможности – недостаточный уровень владения необходимыми методами извлечения такого рода данных. Однако при ближайшем рассмотрении оказывается, что дело не только в гипотетическом нежелании некоторых кросс-культурных исследователей осваивать современные методы работы с данными. Решая одни проблемы, привлечение больших данных сохраняет другие и создает третьи. Рассмотрим их подробнее.

Ограничения кросс-культурных исследований с использованием больших данных

Значительная часть ограничений, связанных с использованием больших данных, обусловлена тем, что, множества «большие данные» и «открытые данные» далеко не всегда совпадают. Это разграничение важно обозначить в связи с тем, что, во многом благодаря легендам вокруг деятельности компании Cambridge Analytica, в массовом сознании начал складываться прототипический образ исследования с использованием больших данных как построения полного профиля человека и предсказания его поведения по цифровому следу его повседневной активности, которой сам этот человек не придает особого значения и не подозревает о том, какую ценную информацию можно извлечь из ее цифровых следов при условии достаточно хорошего алгоритма [5]. Действительно, в ряде случаев существуют весьма сильные корреляции между отдельными характеристиками цифрового следа и индивидуальными характеристиками, что позволяет создавать модели с высокой предсказательной силой. Однако такие сильные связи, из которых можно сконструировать относительно простые, а не многокомпонентные, и при этом эффективные инструменты предиктивной аналитики, зависят не только и не столько от доступа к большому и регулярно обновляющемуся массиву данных, сколько от способности исследователей выбирать из всего многообразия данных те параметры, которые обладают наибольшей предсказательной силой. Иначе говоря, здесь воспроизводятся те же задачи формулирования хороших гипотез, что и при работе с данными другого типа. Разница в том, что в случае неудачного выбора переменных при работе с большими данными достаточно скорректировать алгоритм и нет необходимости заново проводить весьма дорогостоящую процедуру сбора данных. Казалось бы, именно для кросс-культурных исследований это обстоятельство – существенная экономия временных, финансовых и организационных ресурсов – имеет особенно большое значение. Это так. Однако применительно к кросс-культурным исследователям это важное преимущество может быть сведено на нет риском неэквивалентности процессов генерирования однородных по формату и локализации открытых больших данных. Так, пользователи одной и той же социальной сети, даже при условии ее достаточно выраженной популярности в репрезентативном количестве стран, могут различаться по возрасту (в зависимости от того, когда эта социальная сеть стала популярна в данной стране), социальному статусу и уровню образования (в зависимости от того, как именно используются в той или иной стране ресурсы этой социальной сети). Последнее обстоятельство отражает ключевую проблему кросс-культурной эквивалентности: в разных странах одинаковый функционал, предоставляемый техническими ресурсами одной и той же социальной сети, быстро обрастает различными социальными нормами поведения. Эти нормы могут регулировать, например, то, какой контент уместен в этой сети – например, рассматривается ли она как пространство для развлечения, делового общения, серьезного обсуждения социально значимых вопросов или поддержания личных контактов с географически удаленными внесетевыми знакомыми. Социально-демографическая неэквивалентность приводит к тому, что однотипный контент генерируется в разных культурах разными катего-

риями населения, содержательная неэквивалентность – к тому, что сопоставимый контент для наполнения отдельных переменных в некоторых культурах вообще не генерируется. Особенно осторожно в связи с этими различиями следует относиться к открытым большим данным в транснациональных материалах социальных сетей (например, комментарии к новостным материалам СМИ с международной аудиторией): и социально-демографические, и социально-нормативные различия там присутствуют, однако в еще менее явном виде. Кроме того, интерес к международным СМИ и достаточные ресурсы для их потребления и участия в обсуждении материалов (например, знание иностранных языков) в разных культурах также наблюдаются у представителей различных, неэквивалентных социально-демографических категорий. Таким образом, применение открытых больших данных, многократно увеличивая объем доступных данных, также амплифицирует проблему неэквивалентности.

Обращение к большим данным, которые не относятся к категории открытых, отчасти снимает проблему неэквивалентности контента. Применительно к закрытым персональным данным чаще, чем при работе с открытыми данными, речь идет об однотипных специфических паттернах поведения. Например, клиенты интернет-магазина, пользователи определенного мобильного приложения, игроки в определенную компьютерную игру совершают действия из заданного набора, что обусловлено и характером активности, и форматом, предусмотренным конкретным цифровым программным продуктом. Кроме того, в ряде случаев, особенно если идет речь о популярной компьютерной игре, формируются нормы и правила поведения, общие для всех активных пользователей, что делает данные, которые генерируются представителями разных культур, более сопоставимыми. Вместе с тем, у такой большей унифицированности есть и обратная сторона: тем самым резко ограничиваются возможные предметные области исследования. Если, например, пользователи социальных сетей, пусть с определенными оговорками, но все же воспроизводят модели поведения в довольно широком диапазоне ситуаций, то, например, данные о поведении клиентов интернет-магазина безусловно подходят для изучения поведения клиентов интернет-магазинов определенного профиля, менее очевидно – для изучения клиентов интернет-магазинов как таковых, еще менее очевидно – для изучения поведения потребителей вообще и совсем неочевидно – для изучения более общих закономерностей, например, факторов принятия решений такие ограничения могут иметь двойные последствия. С одной стороны, изучение вопросов более общего характера будет чаще рассматриваться на материале, наиболее доступном для изучения, особенно с учетом того, что именно этот материал – поведение, оставляющее цифровой след, – зачастую воспринимается как наиболее точно отражающий новые социальные и технологические тренды и, следовательно, наиболее интересный для изучения, что, однако, вовсе не делает его репрезентативным для всех прочих проявлений универсальных феноменов. С другой стороны, выбор более узко очерченных явлений может при использовании больших данных еще более выражено, чем при специально организуемом сборе данных для исследования, диктоваться практическими соображениями доступности. Наконец, применительно к большим данным, не относящимся к категории открытых, проблема неэквивалентности возникает в связи с тем, что население разных стран может существенно различаться по отношению к угрозам кибербезопасности и надежности операций, совершаемых и/или фиксируемых с использованием цифровых технологий. Соответственно, будут различаться не только частота и характер совершаемых операций, но и их мотивы, что важно и при этом весьма непросто учитывать при интерпретации результатов анализа таких данных.

Заключение

Предпринятое в данной статье рассмотрение возможностей и ограничений использования больших данных в кросс-культурных исследованиях указывает на то, что большие данные, как, пожалуй, любые другие новые методологические инструменты и подходы до них (из относительно недавних можно вспомнить, например, байесовскую статистику), предо-

ставляет некоторые новые возможности и позволяет более эффективно решать ряд проблем, но не представляет собой универсального решения всех методологических и организационных трудностей и, в свою очередь, налагает новые ограничения.

На первый взгляд, эти ограничения преимущественно связаны с тем, что большие данные преимущественно позволяют изучать поведение, оставляющее цифровой след, и с тем, что это поведение в очень разной степени представлено в различных культурах и для разных содержательных контекстов. Это ограничение может уйти в будущем по мере того, как использование цифровых технологий будет становиться все более универсальным и, соответственно, репрезентативным для различных паттернов поведения, в том числе в кросс-культурной перспективе. Однако за этим ограничением скрывается более глубокое различие между кросс-культурными исследованиями с использованием больших данных и данных, специально сгенерированных в исследовательских целях. Это различие связано с одним из лейтмотивов в социальных науках последних десятилетий – темой социального доверия и связанного с ним доверия к информации. В случае специально организованного сбора данных как этапа исследования существует проблема доверия к данным – их аутентичности (особенно если речь идет о кросс-культурном исследовании, где сбор данных в каждой стране осуществляет своя организация: возникает типичный случай проблемы принципала-агента), конструктивной и особенно экологической валидности. В последнее время опасения относительно недостаточной экологической валидности активизировались в форме обсуждения проблемы искренности участников массовых опросов общественного мнения. В случае больших данных, напротив, вопрос доверия к данным как бы решается сам собой: ведь речь идет о следах действий, которые осуществляются естественным образом, без привязки к участию в исследовании. Однако доверие к данным не означает доверия к результатам. Для того, чтобы это трансформация произошла, необходимо доверие к методу и процедуре. Сейчас это доверие, в отличие от доверия к данным, во многом неререфлективное: оно основано отчасти на мифах о сверхэффективных алгоритмах, отчасти – на общем престиже цифровых технологий вообще и машинного обучения и других вариантов искусственного интеллекта в особенности. В дальнейшем на смену этому мифу в массовом сознании придет осознание того, что алгоритм машинного обучения с большой предсказательной силой требует значительных усилий и даже в случае приложения этих усилий может быть разработан только при определенных условиях. Таким образом, на смену проблеме доверия к данным в кросс-культурных исследованиях с использованием больших данных приходит проблема доверия к самому исследованию – с одной стороны, к достаточной компетентности исследователей, с другой – к этичности использования полученных ими результатов. Оба этих направления потенциального недоверия могут налагаться на различные измерения межкультурного неравенства в доступе к новейшим технологиям создания таких алгоритмов и обострить кросс-культурные противоречия в организации кросс-культурного исследования.

Список источников

1. Smith P.B., Bond M.H. Social psychology: Across cultures. Boston : Allyn & Bacon, 1999. 401 p.
2. Wutich A., Brewis A. Data collection in cross-cultural ethnographic research // *Field Methods*. 2019. Vol. 31, № 2. P. 181–189.
3. Фабрикант М.С. Применение нарративного анализа в исследовании больших социальных групп // *Психологический журнал*. 2010. № 1. С. 64–68.
4. Фишер Р. Прикладная кросс-культурная психология социальных изменений и развития общества: на примере исследования благополучия и коррупции // *Психология. Журнал Высшей школы экономики*. 2012. Т. 9, № 2. С. 23–42.
5. Richterich A. How data-driven research fuelled the Cambridge Analytica controversy // *Partecipazione e conflitto*. 2018. Vol. 11, № 2. P. 528–543.

BIG DATA IN CROSS-CULTURAL RESEARCH: TRUST IN INFORMATION AND TRUST IN RESULTS

M.S. Fabrykant

Belarusian State University, Minsk, Belarus

Abstract. The article considers the opportunities and limitations of using big data in cross-cultural research. It discusses how the use of big data allows solving a number of methodological and procedural problems characteristic of cross-cultural research and what new difficulties arise as unintended consequences. A general conclusion is made that the choice between specially generated and big data reflects different aspects of the problem of trust in research: in the former case, the issue of trust in the data itself comes to the fore; in the latter, it is the issue of trust in the results of the study. The problem of trust in the results will increase as machine learning algorithms are demythologized, which, combined with the dissemination of information about cybersecurity threats, can lead, among other things, to aggravation of intercultural contradictions in the organization of cross-cultural research.

Keywords: big data, cross-cultural research, digital footprint, social trust, trust in information

ПРОФЕССИОНАЛЬНЫЕ ОРИЕНТИРЫ СЕЛЬСКОЙ МОЛОДЁЖИ КАК ФАКТОР МИГРАЦИИ В ГОРОД

О.В. Ярмак, А.Ю. Ткаченко

Севастопольский государственный университет, Севастополь, Россия

Аннотация. В статье рассматриваются результаты социологического исследования профессиональных ориентиров сельской молодёжи как фактора миграции в город. В современной России наблюдается тенденция массового миграционного оттока сельской молодёжи в города. Данная тенденция негативно сказывается на развитии сельской местности, государственной экономике, культуре, демографической ситуации. Одной из причин миграции молодёжи из сёл в города являются профессиональные ориентиры молодых людей. Проведя социологическое исследование и рассмотрев профессиональные ориентиры сельской молодёжи как фактор миграции, авторы сделали вывод, что стремление сельской молодёжи переезжать и жить в городах имеет социально-экономические причины: качество жизни и условия для профессионального и личностного роста в городах выше, чем в сельской местности. Сельская молодёжь, в первую очередь, ориентируется на определённый финансовый уровень.

Ключевые слова: человеческий капитал, профессиональные ориентиры, сельская молодёжь, миграция, факторы миграции

Финансовая поддержка: исследование выполнено при поддержке РФФИ в рамках научного проекта № 19-29-07443 на тему «Научно-образовательные центры как фактор формирования человеческого капитала России: формат создания научно-образовательных центров мирового уровня согласно Указу Президента «О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года».

Сельская молодёжь является одним из ключевых элементов в формировании рабочей силы в сельской местности. В условиях современной России можно наблюдать тенденцию оттока молодых людей из села в городскую местность. Такое развитие событий ставит под угрозу возможности обновления и пополнения материального и интеллектуального потенциала российского села. Важно понимать, что экономически стабильные и социально развитые сёла – залог стабильности, независимости и продовольственной безопасности государства, поэтому развитие сёл должно стать приоритетным направлением государственной политики.

Одной из причин миграции сельской молодёжи в город являются профессиональные ориентиры. С течением времени социально-экономическое устройство в стране изменяется. Вместе с ним модифицируются и молодёжные профессиональные ориентиры. Следовательно, профессиональные ориентиры нынешней молодёжи весьма отличаются от тех, которые были свойственны прошлым поколениям.

Актуальность данной темы обусловлена уровнем социального положения сельской молодёжи в государстве. Особенностью сельской молодёжи как социальной группы является то, что её планы и стратегии, а также ориентиры определяют будущее всей страны. Таким образом, рассмотрев профессиональные ориентиры как фактор миграции, можно найти пути к решению проблемы, существующей в реалиях современного российского села.

Основная проблема состоит в том, что массовый миграционный отток молодёжи в города ставит под угрозу развитие сельской местности. Это, в свою очередь, наносит удар по государственной экономике, культуре, демографической ситуации. Сельская местность остаётся без высококвалифицированных специалистов, а уровень его развития может достигнуть кризисного уровня.

Данная проблематика прослеживается в нескольких социологических работах современных авторов. Так, отечественный историк Е.Д. Вознесенская в своей работе отмечает, что сельская молодёжь не ориентирована на смену профессионального и социального статуса [1]. В её среде преобладают потребности в гарантированном минимуме стабильности, социальных льготах и справедливой оплате труда. В формировании профессиональных траекторий сельской молодёжи образование не рассматривается как главный механизм социального восхождения. В социальном плане переезд из села в город – шаг «наверх».

Политологи И.Е. Абрамова и Т.В. Пономаренко делают акцент на том, что для сельской молодёжи свойственна слабая конкурентоспособность по сравнению с другими группами трудоспособного населения [2]. Основное направление реформирования социально-трудовой сферы села и молодёжной политики лежит в сфере роста заработной платы и повышения доходности сельскохозяйственного труда. Существует неудовлетворённость тем, что работа не соответствует специальности. Это косвенно свидетельствует о неудовлетворительности системы профессиональной подготовки молодежи. Еще одним фактором, играющим ключевую роль в молодёжной политике, является жилищный вопрос.

Существует ряд факторов, которые влияют на выбор профессиональных ориентиров индивида. Исследователь Н.А. Грачёва выделяет пять основных групп факторов выбора профессиональных ориентиров: социально-экономические и демографические, инфраструктурно-географические, имиджевые, правовые, личностные и психологические [3]. Эти факторы также создают у человека материальные и идеалистические установки. Выбор профессиональных ориентиров – это соотнесение приобретенных жизненных установок с возможностями их воплощения в жизнь.

Для полного раскрытия темы работы Центром социологических исследований Севастопольского государственного университета было проведено социологическое исследование по смешанной стратегии методами опроса, фокус-группового интервью и контент-анализа.

В опросе респондентами выступали студенты первого курса, которые являются уроженцами сёл. Так как исследование носило пилотажный характер, то объём выборки составил 22 человека, из которых 16 женщин и 6 мужчин в возрасте от 18 до 19 лет. Для авторов являлось важным опросить данную категорию студентов, так как у них сформировалось представление о своём профессиональном будущем относительно недавно. Соответственно, они уже имеют общие представления о жизни и работе в городе, но при этом ещё не полностью приняли и социализировались в городской среде, нежели учащиеся старших курсов.

Больше половины респондентов определились с будущей профессией (64%), и это почти в два раза больше, чем доля неопределившихся (36%). Половина опрошенных частично удовлетворена выбором своей профессии. Наиболее распространёнными причинами выбора профессии оказались востребованность профессии и соответствие профессии личным способностям, знаниям и возможностям. Самые значимые критерии в студенческой среде при выборе будущей работы – это высокая заработная плата и получение удовольствия от работы. Большинство опрошенных не безразлично своё место работы (86%). 60% респондентов не планируют заниматься предпринимательской деятельностью, но 40% респондентов планирует стать предпринимателями. Подавляющее количество опрошенных не рассматривают трудоустройство на производстве (91%).

Выяснилась интересная тенденция: большинство не планируют трудоустроиваться по специальности в сельской местности. При этом наличие собственного жилья является самым значимым фактором, способствующим трудоустройству в сельской местности и возврату молодого специалиста в село.

Наиболее популярными причинами, по которым опрошенные не планируют трудоустроиваться в сельской местности, являются низкий уровень заработной платы, ограниченный выбор профессий, отсутствие возможности самореализации.

Половина респондентов (54%) всё же говорит, что в их сельской местности есть возможность трудоустройства по специальности.

Анализируя полученные данные, можно отметить, что существуют две группы студентов из сельской местности: те, которые хотят трудиться по специальности в сельской местности за предоставление жилья, и те, которые не готовы возвращаться в село после обучения, даже за достаточное денежное вознаграждение.

Таким образом, по результатам авторского исследования можно сделать вывод, что профессиональные ориентиры сельской молодёжи формируются, в первую очередь, под влиянием экономических и личностных факторов. При выборе профессии сельская молодёжь

ориентируется на высокий заработок и получение удовольствия от работы. В сельской местности планируют работать те, у кого имеется собственное жильё или кто его получит. Главными причинами отказа работать в сельской местности являются низкая заработная плата, ограниченный выбор профессий и отсутствие возможности самореализации.

Для того, чтобы выяснить, каким образом профессиональные ориентиры сельской молодёжи влияют на миграцию в город, в рамках исследования были проведены четыре фокус-групповых интервью со студентами СевГУ (8 женщин и 4 мужчины из сельской местности в возрасте от 19 до 22 лет).

Почти все участники фокус-групп (92%) согласились с тем, что в основном сельская молодёжь мигрирует в города для получения образования и из-за стремления улучшить качество жизни. Они отметили, что «в их сёлах нет никаких учебных заведений, кроме школ, поэтому переезд молодых людей из села в город – это логичный шаг, который является необходимостью, если человек настроен на лучшую жизнь». Было высказано мнение о том, что жилищные условия в сельской местности лучше и жильё намного дешевле; о том, что в сельской местности жизнь более безопасна; о том, что продолжительность жизни людей в сёлах выше, чем в городе, за счёт здорового, качественного питания и отсутствия того уровня загрязнения окружающей среды, который наблюдается в городе.

Участниками фокус-групп были названы следующие причины переезда из села в город: потребность в получении высшего образования; высокий уровень качества жизни в городе; большой спектр возможностей для саморазвития в городе; большое количество знакомств в городе; однообразный досуг в селе; плохо развитая инфраструктура в селе; желание уехать от родителей. Все участники отметили, что переезд из села в город – это лучший вариант для молодого человека, так как город предоставляет намного больше социальных возможностей.

75% участников фокус-групп заявили, что не планируют трудоустроиться в сельской местности по специальности. Один участник планирует поработать в селе, но только для получения стажа с дальнейшим переездом в город, а один планирует вернуться в село на работу по семейным обстоятельствам.

Основными причинами, из-за которых участники фокус-групп не планируют трудоустроиться в селе, оказываются отсутствие возможности трудоустройства по специальности, низкий уровень заработной платы и неудовлетворительные условия жизни.

Около 60% участников фокус-групп сказали, что у них нет возможности трудоустроиться по специальности в той сельской местности, в которой они проживали. Отмечается, что рабочих мест недостаточно, поэтому шанс получения должности крайне низок. Также было высказано, что трудоустроиться можно, но не по специальности. Кроме этого, было высказано мнение об обесценивании труда молодых специалистов в селе.

Почти все участники (92%) заявили, что готовы будут трудоустроиться в селе за высокую заработную плату; 17% готовы работать за предоставление жилья. В качестве условий, которые смогли бы послужить стимулом к работе в селе, можно обозначить территориально близкую расположенность села к городу, наличие хорошо развитой инфраструктуры в селе, наличие рабочих мест по специальности, снижение пенсионного возраста для людей, работающих в сельской местности.

По результатам проведённого исследования можно сделать вывод, что сельская молодёжь мигрирует из села в город ради получения образования и улучшения качества жизни. В связи с отсутствием возможности трудоустройства, низким уровнем заработной платы и неудовлетворительными условиями жизни в селе молодые люди не планируют возвращаться в сельскую местность для дальнейшего трудоустройства и проживания. Сельская молодёжь готова возвращаться в село на работу на условиях высокого финансового обеспечения, близости села к городу и развитой инфраструктуры непосредственно в селе.

Стремление сельской молодёжи переезжать и жить в городах обусловлено социально-экономическими обстоятельствами: качество жизни и условия для профессионального и личностного роста в городах выше, чем в сельской местности. Сельская молодёжь, в первую

очередь, ориентируется на определённый финансовый уровень. Таким образом, для того, чтобы сделать сельскую местность привлекательной для молодых людей, необходимо создать условия для экономического, социального и культурного развития села, необходимо обеспечить молодым специалистам высокий уровень заработной платы, предоставить возможность для профессионального развития, увеличить количество рабочих мест, создать программу по предоставлению жилья молодым специалистам, улучшать инфраструктуру в сельской местности.

Список источников

1. Вознесенская Е.Д. Молодые рабочие. Из села в город // Социологический журнал. 2013. № 3. С. 75–91.
2. Абрамова И.Е., Пономаренко Т.В. Формирование жизненных стратегий сельской молодежи: тенденции и возможности государственного влияния // Государственное и муниципальное управление. Ученые записки. 2012. № 4. С. 1–6.
3. Грачёва Н.А., Профорентация и миграция молодежи как важнейшие элементы развития регионального рынка труда // Уровень жизни населения регионов России. 2016. № 3 (201). С. 4–5.

PROFESSIONAL GUIDELINES OF RURAL YOUTH AS A FACTOR OF MIGRATION TO THE CITY

O.V. Yarmak, A.Yu. Tkachenko

Sevastopol State University, Sevastopol, Russian Federation

Abstract. The article deals with the results of a sociological study of professional orientations of rural youth as a factor of migration to the city. In modern Russia, there is a trend of mass migration outflow of rural youth to cities. This trend has a negative impact on the development of rural areas, the state economy, culture, and demographic situation. One of the reasons for the migration of young people from villages to cities is young people's professional orientation. After conducting a sociological study and considering the professional orientations of rural youth as a factor in migration, the authors conclude that the desire of rural youth to move and live in cities has socioeconomic reasons: the quality of life and conditions for professional and personal growth in cities are higher than in rural areas. Rural youth, first of all, is guided by a certain financial level.

Keywords: human capital, professional guidelines, rural youth, migration, migration factors

СВЕДЕНИЯ ОБ АВТОРАХ

- Александрова Юлия Константиновна** – младший научный сотрудник Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: jalexandrova@data.tsu.ru
- Баймурзина Гузель Римовна** – кандидат экономических наук, главный специалист Научной лаборатории социальных и демографических исследований Башкирского государственного университета (Уфа, Россия). E-mail: guzrim@mail.ru. ORCID 0000-0002-1844-2689.
- Барило Илья Иванович** – кандидат технических наук, доцент кафедры информационных систем и технологий Костромского государственного университета (Кострома, Россия). E-mail: i_barilo@ksu.edu.ru
- Барышев Алексей Андреевич** – кандидат экономических наук, доцент кафедры социологии, старший научный сотрудник Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: barishevnp@mail.ru
- Басина Полина Александровна** – аналитик Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: basina@data.tsu.ru
- Батурина Ольга Андреевна** – кандидат экономических наук, доцент, директор департамента цифрового развития Владивостокского государственного университета экономики и сервиса (Владивосток, Россия). E-mail: Olga.Ruban@vvsu.ru
- Беляева Екатерина Сергеевна** – преподаватель кафедры отраслевой и прикладной социологии Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского (Нижний Новгород, Россия). E-mail: belyaeva_ekaterinka@list.ru
- Благинин Алексей Леонидович** – лаборант Центра прикладного анализа больших данных, магистрант Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: alex-b.l@yandex.ru
- Бушкова-Шиклина Эльвира Васильевна** – кандидат социологических наук, доцент кафедры культурологии, социологии и философии Вятского государственного университета (Киров, Россия). E-mail: kaf_ksif@vyatsu.ru
- Галкин Станислав Алексеевич** – кандидат медицинских наук, младший научный сотрудник НИИ психического здоровья, аналитик Лаборатории экспериментальных методов в общественных и когнитивных науках Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: s01091994@yandex.ru
- Гмарь Дмитрий Викторович** – руководитель Центра информационно-технического обеспечения Владивостокского государственного университета экономики и сервиса (Владивосток, Россия). E-mail: dmitriy.gmar@vvsu.ru
- Гмызина Эльвира Викторовна** – кандидат культурологии, доцент кафедры культурологии, социологии и философии Вятского государственного университета (Киров, Россия). E-mail: elvig22@mail.ru
- Гойко Вячеслав Леонидович** – директор Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: goiko@data.tsu.ru
- Голенок Марина Петровна** – кандидат культурологии, заведующий кафедрой культурологии, социологии и философии Вятского государственного университета (Киров, Россия). E-mail: kaf_ksif@vyatsu.ru

- Городилова Татьяна Сергеевна** – кандидат философских наук, доцент кафедры культурологии, социологии и философии Вятского государственного университета (Киров, Россия). E-mail: gtspost@yandex.ru
- Губанов Александр Валерьевич** – лаборант Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: derzhiarbuz@gmail.com
- Завалин Георгий Сергеевич** – специалист отдела учебной аналитики, магистрант Владивостокского государственного университета экономики и сервиса (Владивосток, Россия). E-mail: Georgiy.Zavalin08@vvsu.ru
- Иванова Татьяна Николаевна** – доктор социологических наук, доцент, профессор, заведующий кафедрой социологии Тольяттинского государственного университета (Тольятти, Россия). E-mail: IvanovaT2005@tltsu.ru
- Исакова Инна Александровна** – кандидат социологических наук, доцент, доцент кафедры общей социологии и социальной работы факультета социальных наук Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского (Нижний Новгород, Россия). E-mail: inna_isakova@mail.ru
- Кашпур Валерия Викторовна** – кандидат филологических наук, доцент кафедры романо-германской филологии Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: valkashpur@inbox.ru
- Корзун Дмитрий Жоржевич** – кандидат физико-математических наук, доцент кафедры информатики и математического обеспечения института математики и информационных технологий, ведущий научный сотрудник, заведующий лабораторией интеллектуального анализа данных Петрозаводского государственного университета (ПетрГУ), заместитель директора по науке центра искусственно-го интеллекта (Петрозаводск, Россия). E-mail: dkorzun@cs.karelia.ru
- Космина Елена Алексеевна** – кандидат педагогических наук, доцент кафедры теории и методики массовой физкультурно-оздоровительной работы Национального государственного Университета физической культуры, спорта и здоровья имени П.Ф. Лесгафта (Санкт-Петербург, Россия). E-mail: E.kosmina@lesgaft.spb.ru
- Крюков Владимир Васильевич** – доктор экономических наук, профессор, проректор по цифровому развитию Владивостокского государственного университета экономики и сервиса (Владивосток, Россия). E-mail: vladimir.kryukov@vvsu.ru
- Кузнецова Наталия Михайловна** – аспирант, старший преподаватель кафедры социальной безопасности и гуманитарных технологий Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского (Нижний Новгород, Россия). E-mail: nat-09@yandex.ru
- Кушова Ирина Андреевна** – кандидат философских наук, доцент кафедры культурологии, социологии и философии Вятского государственного университета (Киров, Россия). E-mail: kaf_ksif@vyatsu.ru
- Лагутин Юрий Викторович** – ассистент Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского (Нижний Новгород, Россия). E-mail: lagutinyury@gmail.com
- Лакман Ирина Александровна** – кандидат технических наук, заведующий лабораторией исследования социально-экономических проблем регионов Башкирского государственного университета (Уфа, Россия). E-mail: Lackmania@mail.ru. ORCID: 0000-0001-9876-9202.
- Митягина Екатерина Владимировна** – доктор социологических наук, доцент, проректор по развитию на основе анализа данных Вятского государственного университета (Киров, Россия). E-mail: mityagina2004@yandex.ru

- Мищенко Екатерина Сергеевна** – лаборант Лаборатории экспериментальных методов в общественных и когнитивных науках Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: ekaterina.mish@data.tsu.ru
- Москвин Артем Сергеевич** – кандидат культурологии, доцент кафедры культурологии, социологии и философии Вятского государственного университета (Киров, Россия). E-mail: kaf_ksif@vyatsu.ru
- Мосунова Людмила Александровна** – доктор психологических наук, доцент, профессор кафедры интегрированных коммуникаций Вятского государственного университета (Киров, Россия). E-mail: lmosunova@hotmail.com
- Найденко Дарья Григорьевна** – аналитик Лаборатории экспериментальных методов в общественных и когнитивных науках Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: naidenko@data.tsu.ru
- Новичкова Анастасия Сергеевна** – студент кафедры социологии Тольяттинского государственного университета (Тольятти, Россия). E-mail: oblagko2000@mail.ru
- Новожилова Анастасия Евгеньевна** – аналитик Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: bry-anastasiya@yandex.ru
- Одегова Ксения Игоревна** – старший лаборант кафедры культурологии, социологии и философии Вятского государственного университета (Киров, Россия). E-mail: kaf_ksif@vyatsu.ru
- Петрищева Мария Алексеевна** – студентка факультета журналистики Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: goodmashik@mail.ru
- Петров Евгений Юрьевич** – лаборант Суперкомпьютерного центра, аспирант Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: petrov@data.tsu.ru
- Петрова Ирина Эдуардовна** – заведующий кафедрой отраслевой и прикладной социологии Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского (Нижний Новгород, Россия). E-mail: irinapetrovay@yandex.ru
- Сайфулин Эдуард Риялович** – техник Суперкомпьютерного центра Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: sayfulin@data.tsu.ru
- Саркисова Анна Юрьевна** – кандидат филологических наук, доцент, младший научный сотрудник Центра прикладного анализа больших данных Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: sarkisova@data.tsu.ru
- Сербина Галина Николаевна** – заместитель директора Научной библиотеки Национального исследовательского Томского государственного университета (Томск, Россия). E-mail: gnserbina@yandex.ru
- Ситникова Ирина Викторовна** – ведущий социолог учебно-исследовательской лаборатории социологических исследований, преподаватель кафедры отраслевой и прикладной социологии факультета социальных наук Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского (Нижний Новгород, Россия). E-mail: april@fsn.unn.ru
- Стефановский Дмитрий Владимирович** – руководитель отделения интеллектуальных систем в гуманитарной сфере Российского государственного гуманитарного университета (Москва, Россия). E-mail: stefanovskij.dv@rggu.ru
- Тимирьянова Венера Маратовна** – доктор экономических наук, заместитель заведующего лабораторией исследования социально-экономических проблем регионов Башкирского государственного университета (Уфа, Россия). E-mail: 79174073127@mail.ru

Ткаченко Александра Юрьевна – студент 3 курса по направлению «Социология» Севастопольского государственного университета (Севастополь, Россия). E-mail: Sasha_Stanton@mail.ru

Фабрикант Маргарита Сауловна – кандидат социологических наук, кандидат психологических наук, доцент, доцент кафедры социальной и организационной психологии факультета философии и социальных наук Белорусского государственного университета (Минск, Беларусь). E-mail: marharyta.fabrylant@gmail.com

Цуранов Эдуард Владимирович – ведущий программист Центра информационно-технического обеспечения Владивостокского государственного университета экономики и сервиса (Владивосток, Россия). E-mail: Eduard.Tsuranov@vvsu.ru

Ярмак Ольга Валерьевна – кандидат социологических наук, доцент, заведующий кафедрой социальных коммуникаций Севастопольского государственного университета (Севастополь, Россия). E-mail: olga_yarmak@inbox.ru

ABOUT THE AUTHORS

Yulia K. Aleksandrova, junior researcher, Center for Applied Big Data Analysis of the National Research Tomsk State University (Tomsk, Russian Federation). E-mail: jalexandrova@data.tsu.ru

Guzel R. Baimurzina, Cand. Sci. (Economics), chief specialist, Scientific Laboratory for Social and Demographic Research, Bashkir State University (Ufa, Russian Federation). E-mail: guzrim@mail.ru

Ilya I. Barilo, Cand. Sci. (Engineering), associate professor, Department of Information Systems and Technologies, Kostroma State University (Kostroma, Russian Federation). E-mail: i_barilo@ksu.edu.ru

Aleksey A. Baryshev, Cand. Sci. (Economics), associate professor, Faculty of Philosophy, National Research Tomsk State University; senior researcher, Center for Applied Big Data Analysis, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: barishevnp@mail.ru

Polina A. Basina, analyst, Center of Applied Big Data Analysis, National Research Tomsk State University (Tomsk, RF) E-mail: basina@data.tsu.ru

Olga A. Baturina, Cand. Sci. (Economics), associate professor, director of the Department of Digital Development, Vladivostok State University of Economics and Service (Vladivostok, Russian Federation). E-mail: Olga.Ruban@vvsu.ru

Ekaterina S. Belyaeva, lecturer, Department of Sectoral and Applied Sociology, Lobachevsky State University of Nizhny Novgorod – National Research University (Nizhny Novgorod, Russian Federation). E-mail: belyaeva_ekaterinka@list.ru

Alexey L. Blagin, laboratory assistant, Center for Applied Big Data Analysis, National Research Tomsk State University; master's student, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: alex-b.l@yandex.ru

Elvira V. Bushkova-Shiklina, Cand. Sci. (Sociology), associate professor, Department of Cultural Studies, Sociology and Philosophy, Vyatka State University (Kirov, Russian Federation). E-mail: kaf_ksif@vyatsu.ru

Marharyta S. Fabrykant, Cand. Sci. (Sociology), Cand. Sci. (Psychology), associate professor, Department of Social and Organizational Psychology, Faculty of Philosophy and Social Sciences, Belarusian State University (Minsk, Republic of Belarus). Email: marharyta.fabrykant@gmail.com

Stanislav A. Galkin, Cand. Sci. (Medical Sciences), junior researcher, Mental Health Research Institute, Tomsk National Research Medical Center of the Russian Academy of Sciences; analyst, Laboratory of Experimental Methods in Social and Cognitive Sciences, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: s01091994@yandex.ru

Dmitry V. Gmar, head of the Information and Technical Support Center, Vladivostok State University of Economics and Service (Vladivostok, Russian Federation). E-mail: dmitriy.gmar@vvsu.ru

Elvira V. Gmyzina, Cand. Sci. (Cultural Studies), associate professor, Department of Cultural Studies, Sociology and Philosophy, Vyatka State University (Kirov, Russian Federation). E-mail: elvig22@mail.ru

Vyacheslav L. Goiko, director of the Center for Applied Big Data Analysis, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: goiko@data.tsu.ru

Marina P. Golenok, Cand. Sci. (Cultural Studies), head of the Department of Cultural Studies, Sociology and Philosophy, Vyatka State University (Kirov, Russian Federation). E-mail: kaf_ksif@vyatsu.ru

Tatyana S. Gorodilova, Cand. Sci. (Philosophy), associate professor, Department of Cultural Studies, Sociology and Philosophy, Vyatka State University (Kirov, Russian Federation). E-mail: gtspost@yandex.ru

Alexander V. Gubanov, laboratory assistant, Center for Applied Big Data Analysis, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: derzhiarbuzz@gmail.com

- Tatiana N. Ivanova**, Dr. Sci. (Sociology), associate professor, professor, head of the Department of Sociology, Togliatti State University (Togliatti, Russian Federation). E-mail: IvanovaT2005@tlttsu.ru
- Inna A. Isakova**, Cand. Sci. (Sociology), docent, associate professor of Department of General Sociology and Social Work, Lobachevsky State University of Nizhny Novgorod – National Research University (Nizhny Novgorod, Russian Federation). E-mail: inna_isakova@mail.ru
- Valeriya V. Kashpur**, Cand. Sci. (Philology), associate professor, Faculty of Philology, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: valkashpur@inbox.ru
- Dmitry Zh. Korzun**, Cand. Sci. (Physics and Mathematics), associate professor, Department of Computer Science, Institute of Mathematics and Information Technology, Petrozavodsk State University; leading researcher, head of the Data Mining Lab, Petrozavodsk State University; vice-director for Research of the Artificial Intelligence Center, Petrozavodsk State University (Petrozavodsk, Russian Federation). E-mail: dkorzun@cs.karelia.ru
- Elena A. Kosmina**, Cand. Sci. (Pedagogy), associate professor, Department of Theory and Methodology of Mass Physical Culture and Wellness Work, Lesgaft National State University of Physical Education, Sport and Health (St. Petersburg, Russian Federation). E-mail: E.kosmina@lesgaft.spb.ru
- Vladimir V. Kryukov**, Dr. Sci. (Economics), professor, vice-rector for Digital Development, Vladivostok State University of Economics and Service (Vladivostok, Russian Federation). E-mail: vladimir.kryukov@vvsu.ru
- Irina A. Kushova**, Cand. Sci. (Philosophy), associate professor, Department of Cultural Studies, Sociology and Philosophy, Vyatka State University (Kirov, Russian Federation). E-mail: kaf_ksif@vyatsu.ru
- Natalia M. Kuznetsova**, postgraduate student, senior lecturer, Department of Social Security and Humanitarian Technologies, Lobachevsky State University of Nizhny Novgorod – National Research University (Nizhny Novgorod, Russian Federation). E-mail: nat-09@yandex.ru
- Irina A. Lackman**, Cand. Sci. (Engineering), head of the Laboratory for the Study of Socio-Economic Problems of Regions, Bashkir State University (Ufa, Russian Federation). E-mail: Lackmania@mail.ru
- Yury V. Lagutin**, assistant, Lobachevsky State University of Nizhny Novgorod – National Research University (Nizhny Novgorod, Russian Federation). E-mail: lagutinyury@gmail.com
- Ekaterina S. Mishchenko**, laboratory assistant, Laboratory of Experimental Methods in Social and Cognitive Sciences, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: ekaterina.mish@data.tsu.ru
- Ekaterina V. Mitiagina**, Dr. Sci. (Sociology), associate professor, vice-rector for Development Based on Data Analysis, Vyatka State University (Kirov, Russian Federation). E-mail: mityagina2004@yandex.ru
- Artem S. Moskvin**, Cand. Sci. (Cultural Studies), associate professor, Department of Cultural Studies, Sociology and Philosophy, Vyatka State University (Kirov, Russian Federation). E-mail: kaf_ksif@vyatsu.ru
- Lyudmila A. Mosunova**, Dr. Sci. (Psychology), docent, professor, Department of Integrated Communications, Vyatka State University (Kirov, Russian Federation). E-mail: lmosunova@hotmail.com
- Daria G. Naidenko**, analyst, Laboratory of Experimental Methods in Social and Cognitive Sciences, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: naidenko@data.tsu.ru
- Anastasia S. Novichkova**, student, Department of Sociology, Togliatti State University (Togliatti, Russian Federation). E-mail: oblagko2000@mail.ru
- Anastasiya E. Novozhilova**, analyst, Center for Applied Big Data Analysis, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: bry-anastasiya@yandex.ru

- Ksenia I. Odegova**, senior assistant, Department of Cultural Studies, Sociology and Philosophy, Vyatka State University (Kirov, Russian Federation). E-mail: kaf_ksif@vyatsu.ru
- Maria A. Petrishcheva**, student, Faculty of Journalism, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: goodmashik@mail.ru.
- Evgeny Yu. Petrov**, laboratory assistant, Supercomputer Center, National Research Tomsk State University; postgraduate student, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: petrov@data.tsu.ru
- Irina E. Petrova**, head of the Department of Sectoral and Applied Sociology, Lobachevsky State University of Nizhny Novgorod – National Research University (Nizhny Novgorod, Russian Federation). E-mail: irinapetrovay@yandex.ru
- Eduard R. Sajfulin**, technician, Supercomputer Center, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: sayfulin@data.tsu.ru
- Anna Yu. Sarkisova**, Cand. Sci. (Philology), docent, junior researcher, Center for Applied Big Data Analysis, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: sarkisova@data.tsu.ru
- Galina N. Serbina**, deputy director, Research Library, National Research Tomsk State University (Tomsk, Russian Federation). E-mail: gnserbina@yandex.ru
- Irina V. Sitnikova**, leading sociologist, Training and Research Laboratory of Sociological Research, Lobachevsky State University of Nizhny Novgorod – National Research University; lecturer, Department of Sectoral and Applied Sociology, Faculty of Social Sciences, Lobachevsky State University of Nizhny Novgorod – National Research University (Nizhny Novgorod, Russian Federation). E-mail: april@fsn.unn.ru
- Dmitry V. Stefanovsky**, head of the Department of Intelligent Systems in the Humanities, Russian State University for the Humanities (Moscow, Russian Federation). E-mail: stefanovskij.dv@rggu.ru
- Venera M. Timiryanova**, Dr. Sci. (Economics), deputy head of the Laboratory for the Study of Socio-Economic Problems of Regions, Bashkir State University (Ufa, Russian Federation). E-mail: 79174073127@mail.ru
- Alexandra Yu. Tkachenko**, sociology student, Sevastopol State University (Sevastopol, Russian Federation). E-mail: Sasha_Stanton@mail.ru
- Eduard V. Tsuranov**, lead programmer, Information Technology Support Center, Vladivostok State University of Economics and Service (Vladivostok, Russian Federation). E-mail: Eduard.Tsuranov@vvsu.ru
- Olga V. Yarmak**, Cand. Sci. (Sociology), associate professor, head of the Department of Social Communications, Sevastopol State University (Sevastopol, Russian Federation). E-mail: olga_yarmak@inbox.ru
- Georgy S. Zavalin**, specialist, Educational Analytics Department, Vladivostok State University of Economics and Service; master's student, Vladivostok State University of Economics and Service (Vladivostok, Russian Federation). E-mail: Georgiy.Zavalin08@vvsu.ru

Научное издание

**БОЛЬШИЕ ДАННЫЕ
И ПРОБЛЕМЫ ОБЩЕСТВА**

**Сборник статей по итогам
Международной научной конференции
(Киров, 19–20 мая 2022 г.)**

Редактор К.Г. Шилько
Оригинал-макет А.И. Лелоюр
Дизайн обложки Л.Д. Кривцовой

Подписано к печати 12.09.2022 г. Формат 60×84¹/₈.

Бумага для офисной техники.

Печ. л. 21. Усл. печ. л. 19,5.

Тираж 500 экз. Заказ № 5095.

Отпечатано на оборудовании
Издательства Томского государственного университета
634050, г. Томск, пр. Ленина, 36
Тел. 8+(382-2)–52-98-49
Сайт: <http://publish.tsu.ru>
E-mail: rio.tsu@mail.ru

ISBN 978-5-907572-19-5



9 785907 572195 >