
ОНЛАЙН-ИССЛЕДОВАНИЯ

С.Г. Давыдов, О.С. Логунова
(Москва)

ПРОЕКТ «ИНДЕКС ЦИФРОВОЙ ГРАМОТНОСТИ»: МЕТОДИЧЕСКИЕ ЭКСПЕРИМЕНТЫ

В рамках проекта «Индекс цифровой грамотности» был проведен ряд методических экспериментов, связанных с анализом «больших данных» и данных, полученных в результате опроса пользователей социальной сети. В результате были зафиксированы существенные расхождения между данными, полученными в опросе ВЦИОМа, и результатами эксперимента. Анализ «больших данных» показал более равномерное распределение поисковой активности и интереса к интернет-СМИ пользователей по федеральным округам. В зависимости от типа исходных данных выделяются различные округа-лидеры и отстающие. Опрос в социальной сети также продемонстрировал существенные отличия в рассмотренных показателях по сравнению с оффлайн-опросом ВЦИОМа – вплоть до прямо противоположных оценок на уровне федеральных округов. В статье обсуждаются возможные причины, обуславливающие эти различия.

Ключевые слова: цифровая грамотность, большие данные, массовый опрос.

Сергей Геннадьевич Давыдов – кандидат философских наук, заместитель декана по учебной работе, доцент факультета коммуникаций, медиа и дизайна, Национальный исследовательский университет «Высшая школа экономики». E-mail: sdavydov@hse.ru.

Ольга Сергеевна Логунова – кандидат социологических наук, доцент департамента социологии факультета социальных наук, Национальный исследовательский университет «Высшая школа экономики». E-mail: ologunova@hse.ru.

Постановка исследовательской задачи

Современный этап развития информационно-коммуникационных технологий сопровождается ростом объемов производимой информации. Структуры, обладающие этой информацией, справедливо рассматривают ее в качестве актива, который может быть использован во благо не только организаций-собственников, но и широкого круга заинтересованных индустриальных игроков. В то же время для профессионального исследовательского сообщества существование подобной информации становится вызовом, поскольку для ее корректного использования необходима серьезная методическая работа. По причинам, обсуждение которых выходит за рамки настоящей статьи, данная работа проводится далеко не всегда.

Статья посвящена результатам двух методических экспериментов, выполненных в рамках исследования «Индекс цифровой грамотности» (ИЦГ)¹. Теоретическую основу построения индекса составила концепция П. Гилстера [3], рассматривавшего цифровую грамотность как включающую в себя медиаграмотность и умение критического восприятия информации; взгляды А. Букхорста, отмечающего, что к понятию цифровой грамотности содержательно близки компьютерная грамотность, интернет-грамотность, медиа- и информационная грамотность [4]; публикации последних лет М. Фишера, Дж. Авила, М. Банзато, А.Картелли, в которых развивается понятие цифровой грамотности (см.: [5–8]).

При разработке индекса авторы ориентировались в наибольшей степени на следующие проекты. Во-первых, модель цифровой грамотности, предложенная Канадским центром цифровой и меди-

¹ Проект ИЦГ был впервые реализован некоммерческим партнерством РОЦИТ совместно с факультетом коммуникаций, медиа и дизайна НИУ ВШЭ в 2015 г. Основной целью создания индекса стало сравнение уровня развития цифровой грамотности в российских регионах (подробнее о проекте и использованной в нем модели см.: [1; 2]).

аграмотности «МедиаСмартс» [9]. Она представляет собой обобщение ряда других разработок, среди которых проекты рабочей группы Цифровой британской медиаграмотности [10], Европейской сети цифровой грамотности «DigEuLit» [11], а также группы специалистов под руководством Генри Дженкинса [12]. В основе данной модели лежат три идеи:

- освоение инноваций в сфере информационно-коммуникационных технологий;
- нацеленность на конструктивное социальное действие;
- развитие критического мышления и креативных способностей человека.

Во-вторых, проект Фонда «Современная Польша» (Fundacja «Nowoczesna Polska») «Цифровое будущее» (см.: [13; 14]).

В-третьих, был учтен опыт отечественных исследований, в частности разработка «Цифровая компетентность» под руководством Г.У. Солдатовой. Основной слоган рассматриваемой разработки автор определяет следующим образом: «От цифровой грамотности к цифровой компетентности» [15]. Этот слоган был предложен в 2011 г. тремя финскими авторами – Л. Иломаки, М. Лаккала и А. Кантосало – и представлен как новый виток развития движения цифровой грамотности [16]. В понимании Г.У. Солдатовой и ее коллег, цифровая компетентность включает в себя четыре компонента – знания, умения, мотивация, ответственность, причем последняя связана с безопасностью [15].

Рассмотрев имеющиеся на данный момент подходы, авторами исследования была предложена 4-компонентная модель цифровой грамотности, основанная на бинарных оппозициях «технико-технологическое/социогуманитарное» и «возможности/угрозы» [1, с. 91]¹. В целом, под цифровой грамотностью авторами проекта понимается вовлеченность индивида в цифровую среду, подкре-

¹ Более детально процесс операционализации понятия изложен в статье А.В. Шарикова «О четырехкомпонентной модели цифровой грамотности» [1].

пленная наличием компетенций в области потребления, поиска и обработки, а также защиты информации.

В соответствии с проведенной операционализацией, для определения индекса цифровой грамотности необходимо было оценить 20 параметров. Большинство параметров – 16 из 20 – были рассчитаны на основании результатов массового опроса взрослых россиян, проведенного ВЦИОМом в сентябре 2015 г. специально для данного проекта¹. Необходимость производства столь значительного объема информации в рамках проекта была обусловлена двумя факторами. С одной стороны, существующие статистические данные по медиа и информационно-коммуникационным технологиям в субъектах Российской Федерации фрагментарны и не обеспечивают полного понимания индустрии массовых коммуникаций в регионах. С другой – к данным, используемым при построении статистических индексов, предъявляются определенные требования, связанные с надежностью, частотой и скоростью обновления, уровнем детализации и т.д.; дополнительные ограничения приводят к сокращению числа доступных для использования показателей. В результате, помимо опроса ВЦИОМа, для конструирования индекса были использованы следующие источники.

1. Данные по охвату стационарного и мобильного Интернета, а также об уровне потребления социальных медиа предоставлены исследовательской компанией *TNS Россия*. В первых двух случаях источником послужило Установочное исследование *TNS Web Index*

¹ В рамках массового опроса была реализована многоступенчатая стратифицированная территориальная случайная выборка, используемая ВЦИОМом для проведения еженедельных омнибусных опросов, репрезентирующих генеральную совокупность – взрослое население России 18+ – по полу, возрасту, уровню образования, типу населенного пункта и федеральным округам. Опрос был проведен в 46 регионах страны и 132 точках опроса. Случайная ошибка выборки равна ±3,4% (при 95% уровне доверительной вероятности). Подробное описание построения выборочной совокупности приводится на сайте ВЦИОМа [17].

в январе–июне 2015 г., проведенное в городах РФ с населением от 100 тыс. жителей, возраст респондентов – 12+. В третьем случае были взяты данные проекта *TNS Web Index* – вся Россия, возраст респондентов 12–64 лет за аналогичный период времени.

2. Количество интернет-СМИ, зарегистрированных в регионе на душу населения, было рассчитано на основании сведений из базы данных СМИ, зафиксированных Роскомнадзором (по состоянию на октябрь 2015 г.), а также сведений Росстата о численности населения федеральных округов РФ по состоянию на 1 января 2015 г.

Учитывая недостаточность релевантных статистических данных, высокую стоимость репрезентативных всероссийских опросов методом личных квартирных интервью, а также известные ограничения данного метода сбора информации, было принято решение провести в рамках проекта два методических эксперимента¹. Их задача заключалась в оценке потенциала различных источников данных в качестве индикаторов уровня цифровой грамотности. В первом случае речь шла о так называемых больших данных, в нашем случае предоставленных компанией *Data-Centric Alliance* (DCA); во втором – об опросе пользователей одной из ведущих социальных сетей, реализованном при поддержке «Одноклассники.ру» (ОК). В обоих кейсах базой для сравнения служили результаты всероссийского массового опроса, проведенного методом личных интервью по месту жительства респондентов. Эксперименты можно считать успешными, если бы 1) показатели, полученные в результате их проведения, были бы статистически близки к соответствующим показателям опроса; 2) при значимых различиях в значениях ряды данных демонстрировали схожие тенденции (регионы с наиболее высокими и наиболее низкими значениями параметров в том и другом случае совпадали).

¹ Собранные в рамках этих экспериментов данные не были использованы при подсчете индекса и были опубликованы в рамках итогового отчета в качестве дополнительных.

Методический эксперимент 1: анализ больших данных DCA

Первым дополнительным источником информации стали «большие данные», предоставленные компанией DCA. Последней была проведена работа по оценке двух показателей: уровень потребления новостной информации и компетентность в области поиска информации в Интернете.

Цель данного эксперимента – сравнить «большие данные» и ретроспективные самооценки поведения респондентов. Авторы исследования исходили из гипотезы, что полученные результаты будут близки по значениям к результатам массового опроса, что позволит увеличить валидность полученных в результате эксперимента данных и детализировать данные внутри федеральных округов по отдельным субъектам. Таким образом, будут существенно снижены затраты на получение статистических показателей по сравнению с опросом.

Вопросы были сформулированы одинаково и для всероссийского опроса, и для сбора «больших данных»¹, а именно:

КАК ЧАСТО ВЫ ОБРАЩАЕТЕСЬ К ИНТЕРНЕТУ КАК К ИСТОЧНИКУ НОВОСТЕЙ О СТРАНЕ И МИРЕ?

1. Каждый/почти каждый день
2. Несколько раз в неделю
3. Раз в неделю и реже, но чаще, чем раз в месяц
4. Раз в месяц и реже
5. Никогда

¹ При использовании «больших данных» вопросы не задавались пользователям – они были адресованы исследователям для получения соответствующих оценок целевых показателей (частота использования Интернета как источника новостей и частота использования поисковых сервисов).

КАК ЧАСТО ВЫ ПОЛЬЗУЕТЕСЬ ПОИСКОВЫМИ СИСТЕМАМИ (ЯНДЕКС, ГУГЛ И ДР.)?

1. Каждый/почти каждый день
2. Несколько раз в неделю
3. Раз в неделю и реже, но чаще, чем раз в месяц
4. Раз в месяц и реже
5. Никогда

При оценке двух указанных выше показателей была использована одна и та же шкала, что впоследствии упростило процедуру сравнительного анализа.

Компания DCA провела работу по оценке доли пользователей сети Интернет в России, которые могли бы выбрать один из 5 вариантов ответа на каждый из двух вопросов.

Для проведения расчетов по каждому показателю были созданы две разные случайные выборки cookie-файлов объемом около 25 млн экземпляров каждая. Поставщиками данных для аналитической системы DCA служат более 35 компаний, в том числе крупные онлайн-медиа. В cookie policy каждого сервиса прописаны условия согласия пользователя.

Для исследования использовались наборы фактов о каждой cookie за период с 29 июля по 14 сентября 2015 г. В обе выборки попали только те cookie, для которых выполняются следующие правила:

- у cookie был хотя бы один факт до 29 июля;
- для cookie известен город проживания (достаточно длинная история фактов за все время ее существования в системе).

Далее схема подсчета выглядела следующим образом. Для первого показателя – уровень потребления новостной информации в Интернете – считалось за данный полуторамесячный период хотя бы одно посещение сайта СМИ. Среди всех данных в DMP были отобраны те файлы cookie, которые в указанный период зафиксировали хоть одно посещение любой страницы из списка 1000 новостных сайтов. Данный список формировался на основе рейтинга СМИ LiveInternet и включает в себя как крупные

всероссийские СМИ, так и большое количество более мелких региональных информационных площадок.

Поскольку варианты ответов, которые можно выбрать по обоим вопросам, совпадают, методики расчета с определенного момента для обоих вопросов совпадали. После генерации выборок данные представлялись в следующем виде:

- уникальный cookie_id;
- город проживания, определенный для этой cookie;
- набор моментов времени, когда был зарегистрирован каждый из фактов (переход из поисковой выдачи или посещение любой страницы сайтов СМИ из списка) в заданном временном окне в 1,5 месяца.

Далее по набору моментов времени выбирался один из 4 вариантов (кроме варианта «Никогда»), количество каждого из вариантов для каждой cookie_id суммировались по городам, города затем соотносились с одним из 85 субъектов РФ (включая Крым и Севастополь). Правила выбора одного из четырех вариантов, кроме «Никогда», формулировались следующим образом.

Каждый/почти каждый день:

- за 1,5 месяца было не менее 14 дней, когда произошло хоть одно релевантное событие (запросы или посещения СМИ);
- максимальный интервал между двумя соседними днями, когда произошло хоть одно такое событие, не более 3 дней.

Несколько раз в неделю:

- за 1,5 месяца было не менее 10 дней, когда произошло хоть одно релевантное событие (запросы или посещения СМИ);
- максимальный интервал между двумя соседними днями, когда произошло хоть одно такое событие, не более 7 дней.

Раз в неделю и реже, но чаще, чем раз в месяц:

- за 1,5 месяца было не менее 4 дней, когда произошло хоть одно релевантное событие (запросы или посещения СМИ);
- максимальный интервал между двумя соседними днями, в которые произошло хоть одно такое событие, не более 15 дней.

Раз в месяц и реже:

– за 1,5 месяца было не менее 1 релевантного факта (запросов или посещений СМИ);

– по сути, все остальные файлы, попавшие в выборку.

Если cookie была отнесена к варианту 1, то во 2-й уже не попадала и т. д. Ответ «Никогда» выбирался несколько иным образом. Бралась другая независимая выборка cookie, у которых определен город проживания, объемом 1 млн штук, затем в этой выборке вычислялась доля cookie, у которых в истории нет ни одного факта о совершении поисковых запросов или нет ни одного посещения сайтов СМИ. Далее это соотношение мы умножили на объем больших выборок (25 млн) и таким образом получили количество предполагаемых ответов «Никогда» также в разбивке по субъектам РФ.

В результате по каждому субъекту было получено 5 чисел, характеризующих прогнозируемое число каждого из ответов. Следующим шагом эти абсолютные значения переводились в относительные по каждому региону, чтобы исключить смещения, связанные с различной степенью покрытия регионов и разной численностью населения. После чего разные субъекты можно сравнивать между собой по этим показателям.

Таким образом, был получен «слепок» реальных посещений новостных сайтов, которые можно сравнить с результатами опроса жителей по федеральным округам.

Различия данных опроса и «больших данных» объясняются, во-первых, разными основаниями оценивания (рис. 1). Как видно на графике, результаты «больших данных» более равномерны, разница наибольшего и наименьшего значения заметно меньше. Кроме того, лидеры по опросу ВЦИОМа и данным ДСА не совпадают. Подтверждаются данные по Центральному, Уральскому, Южному, Сибирскому, Дальневосточному федеральным округам. Однозначным аутсайдером становится Крымский федеральный округ, тогда как лидеры по результатам опроса ВЦИОМ – Северо-Западный и Северо-Кавказский округа занимают более скромные позиции.

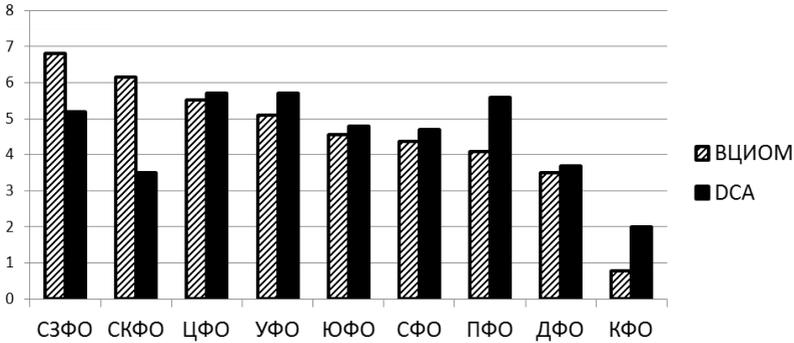


Рис. 1. Сравнение данных опроса ВЦИОМ и больших данных DCA. Показатель «Уровень потребления новостной информации в Интернете»

Кроме того, были получены данные не только с разбивкой на округа, но и по каждому субъекту федерации. Продемонстрируем диспропорции на примере нескольких географических зон. В частности, в Северо-Западном федеральном округе можно обозначить группы активных пользователей, проявляющих интерес к интернет-СМИ на ежедневной основе – это жители Санкт-Петербурга (значение индекса аффинитивности¹ – 146), Калининградской области (значение индекса аффинитивности – 123), Республики Коми (индекс аффинитивности – 115). Отсутствие интереса продемонстрировали жители Ненецкого автономного округа – пользователи данного региона чаще жителей остальных регионов округа попадали в группу тех, кто вообще не использует Интернет для получения новостей. Пользователи Дальневосточного федерального округа достаточно четко разделены в региональной привязке: поисковая активность и интерес к интернет-СМИ характерны для населения

¹ Показатель, отражающий отношение рейтинга по целевой аудитории (в данном случае – доли активных пользователей в Санкт-Петербурге) к рейтингу по базовой аудитории (доли жителей Санкт-Петербурга в выборке).

Приморского и Хабаровского краев. Наименьшая пользовательская активность отличает жителей Чукотского автономного округа.

Во-вторых, дифференциация значений обуславливается разными основами выборки. Для второго показателя – компетентность в области поиска информации в Интернете – в заданный полуторамесячный период отслеживался хотя бы один поисковый запрос.

Исследование поисковой активности основывается на данных переходов из поисковиков на сайты, данные по которым есть в DMP. Иначе говоря, учитывалось не количество поисковых запросов, а реальные переходы на интересующие сайты. Таким образом, данные могут отличаться от результатов исследований, где учитывается охват поисковой выдачи. В нашем случае это исследование объема трафика из поисковых систем. Методика оценки вариантов ответа была построена по аналогии с методикой, описанной выше.

Результаты сравнения данных представлены на *рис. 2*, они также различаются и здесь уже можно проследить несколько закономерностей. Во-первых, заметен высокий уровень рассмотренных компетенций в Северо-Кавказском округе, который вторично не совпадает с «большими данными». Во-вторых, данные ДСА показывают более равномерное распределение, нет ярко выраженного лидера, все восемь округов расположены в небольшом диапазоне. В-третьих, обнаружены совсем другие лидеры, в данном случае это Приволжский и Уральский федеральные округа. Положение остальных, похоже, отличается только отстающими, которыми по результатам анализа «больших данных» становятся Северо-Кавказский округ и Крым.

Одним из основных объяснений становятся стандартные ограничения использования подобных данных. В первую очередь отметим, что полученные соотношения не характеризуют поведение людей в Интернете напрямую. Происходит связывание посещений различных интернет-страниц с реальным поведением пользователей. Кроме того, полученные данные имеют ряд смещений выборки.

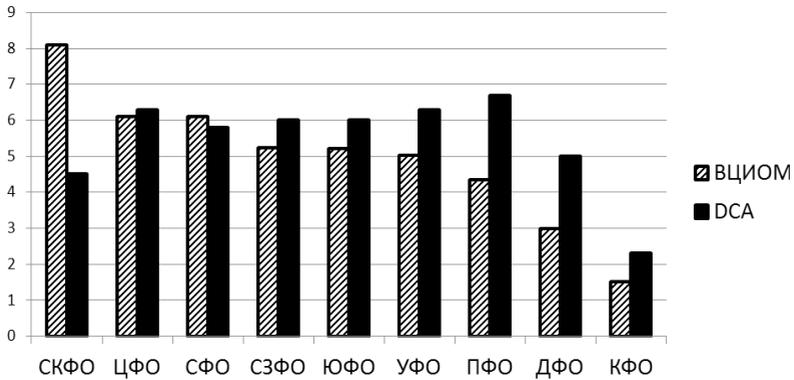


Рис. 2. Сравнение данных опроса ВЦИОМ и больших данных DCA. Показатель «Компетентность в области поиска информации в Интернете»

Значимым оказывается изначальное преобладание группы активных пользователей Интернета, людей среднего возраста. Дополнительно эти люди должны разрешить использование cookie в настройках браузеров. (Если пользователь очистил cookies в браузере/переустановил ОС/поменял жесткий диск/купил новое устройство и т.п., мы больше никогда не получим данных по его старой cookie, с точки зрения системы он станет новым пользователем.) В последнее время постепенно растет число cookie, по которой регистрируются новые факты только на протяжении сравнительного короткого срока. Это связано с использованием приватных режимов в браузерах (например, режим «Инкогнито» в Google Chrome), прокси серверов и прочих маршрутизаторов интернет-трафика. Данные по таким cookie представляются шумом в контексте данного исследования, с которым мы боролись путем введения ограничений на общее число фактов по cookie за всё время и на «распределенность» этих фактов по временной шкале. Таким образом, выборка становится однородной в пользу активных пользователей, дополнительно разрешивших доступ к своим данным.

Третьим ограничением послужил тот факт, что в данном случае не была учтена информация о поведении людей в социальных сетях, т.е. переходы на новостные ресурсы внутри не отслеживались. А такие действия традиционно рассматриваются пользователями как потребление новостной информации в Интернете.

Четвертый фактор – это ситуация, когда один и тот же файл cookie может соответствовать поведению разных пользователей одного и того же ПК, так же как у одного человека может быть несколько cookie на разных устройствах и в разных браузерах. В большинстве случаев выявить эти связи невозможно. Это методическое ограничение активно обсуждается (см.: [18]), однако решение на данном этапе не найдено. Все обозначенные выше ограничения влияют на массив данных и итоговое распределение.

Методический эксперимент 2: опрос в социальной сети Одноклассники

Второй эксперимент заключался в проведение онлайн-опроса среди пользователей социальной сети Одноклассники.ру. Представители последней ссылаются на данные LiveInternet, в соответствии с которыми аудитория ОК в 2015 г. составила 48 млн человек и социально-демографическая структура максимально приближена к структуре населения России.

Исходя из типологии, предложенной А.М. Мавлетовой [19], второй эксперимент был основан на опросе пользователей социальных сетей. Рассмотрим ограничения и возможности данного метода применительно к нашему кейсу. Прежде всего важно отметить ошибки выборки и вопросы, связанные с репрезентативностью результатов. Во-первых, опрашиваемая аудитория ограничена пользователями сети, что значительно сокращает генеральную совокупность. Более того, выборка смещена в сторону активных пользователей социальной сети. Во-вторых, довольно высока доля прерывания опроса, т.е. число анкет, заполненных не

до конца. В данной ситуации нет возможности оказывать давление на респондента, просить заполнить анкету полностью. В-третьих, стихийность выборки – анкету заполняют добровольцы, т.е. используется метод самоотбора. Кроме того, при проведении онлайн опроса сложно проконтролировать добросовестность заполнения анкеты, другими словами, возможность верификации данных практически отсутствует.

Вместе с тем у данного метода есть определенные достоинства. Прежде всего, это скорость опроса – за восемь дней удалось собрать более 16 тыс. анкет по всем субъектам федерации. Во-вторых, существуют возможности конструирования целевой выборки по определенному профилю, в нашем случае это географический таргетинг. Кроме того, исследователи могут сегментировать аудиторию по полу, возрасту, образованию и даже политическим предпочтениям. Еще один значимый плюс – это довольно высокий уровень доверия, отсутствие интервьюера, что минимизирует появление социально одобряемых ответов. Кроме того, существует возможность оперативного контроля, реагирования и доступна коммуникация с респондентами в дальнейшем, что дает возможность уточнить некоторые вопросы или пригласить к дальнейшему обсуждению.

Как и в первом случае, задача эксперимента состояла в том, чтобы сравнить полученные данные с результатами опроса ВЦИОМа на уровне федеральных округов. Дополнительно была получена более высокая детализация ответов – на уровне субъектов федерации.

Для каждого субъекта РФ от имени РОЦИТ в ОК были созданы отдельные опросы. Для соблюдения условий рекрутирования респондентов, каждый опрос показывался на аудиторию выбранного региона. Было два условия участия в опросе – это желание пользователя и установленная географическая локация, которая в дальнейшем позволяла отнести пользователя к определенному субъекту РФ. Контроль за отображением только выбранной географии осуществлялся посредством рекламной сети ОК «My target»,

где контролировалась география показа, частота и количество показов объявлений на отдельного уникального пользователя сети.

Функциональность анкеты была ограничена возможностями опросного инструментария в социальной сети. Анкета состояла из семи вопросов и повторяла формулировки всероссийского опроса. С точки зрения графического отображения интерфейса опрос выглядел как отдельный «медиатоппик», который открывался в дополнительном окне поверх основного интерфейса ОК, и где все вопросы размещались один под другим. Иными словами, для прохождения опроса нужно было выбрать вариант ответа и «скролить» до следующего вопроса. Это и привело к повышенному уровню отказов и высокому уровню незаполненных анкет. Данный показатель заметно различается между регионами.

Всего в опросе приняли участие 16304 человека в период с 30 сентября по 8 октября 2015 г. Была осуществлена невероятная выборка, возраст пользователей – 18+. Не удалось выдержать жесткого квотирования количества участников по регионам, в результате не все субъекты оказались достаточно представлены количественно, и некоторые были исключены из анализа.

Сопоставим распределение ответов на вопрос «Как вы поступите в случае, если будет существовать угроза данным на Вашем электронном устройстве (компьютер, планшет и др.) или оно будет атаковано вирусами?», заданный в рамках двух исследований (*табл. 1*). Отметим, что в рамках опроса ВЦИОМ была использована дополнительная альтернатива «Затрудняюсь ответить», которая во время интервью не зачитывалась. Данный ответ был выбран в 25,0% случаев. Если в случае опроса ВЦИОМ в выборку вошли респонденты, не пользующиеся или редко пользующиеся компьютером и Интернетом, то в опросе ОК преобладали активные пользователи.

Данные представленные в *табл. 1*, свидетельствуют о существенном расхождении результатов, полученных в рамках двух опросов. Действительно, из 27 показателей по 8 зафиксировано

расхождение более чем на 10 процентных пунктов, причем по трем (показатели 2 и 3 ЮФО, показатель 3 ПФО) – более чем на 25 процентных пунктов. Отсюда различия между рангами федеральных округом по каждому из трех параметров (*табл. 2*). Наиболее яркий пример расхождения: по доле тех, кто готов обратиться за помощью к профессионалам, согласно данным ВЦИОМ, КФО занимает первое место, тогда как по данным ОК – последнее.

Выскажем некоторые соображения о природе зафиксированных расхождений. Проводя исследование на определенной онлайн площадке, мы заведомо получаем смещение в сторону как активных пользователей Интернета в целом, так и – если не будут предприняты специальные меры – активных пользователей данной площадки. Сформированная таким образом выборка может приближаться по некоторым параметрам (например, социально-демографическим) к структуре населения, однако по другим параметрам мы неизбежно получим отклонения. В числе последних с высокой долей вероятности окажутся параметры, связанные с медийным потреблением, опытом, поведением и т.д. В связи с этим представляется, что для качественного формирования выборок на базе одной медийной платформы необходимы глубокое понимание специфики аудитории этой платформы по широкому кругу параметров, которые могут оказать влияние на результаты исследования, а также возможности автоматического или анкетного скрининга пользователей для участия в опросе.

Заключение

Несмотря на то что в случае обоих экспериментов требуемого качества данных достичь не удалось (очевидно, что первый кейс с большими данными в партнерстве с DCA представляется более перспективным), подобные методы имеют определенный потенциал. При условии продолжения методической работы и

Таблица 1
ПОВЕДЕНИЕ ЛЮДЕЙ В СЛУЧАЕ ВОЗНИКНОВЕНИЯ УГРОЗЫ ДАННЫМ
НА ИХ ЭЛЕКТРОННОМ УСТРОЙСТВЕ

	ЦФО	СЗФО	ЮФО	СКФО	ПФО	УФО	СФО	ДФО	КФО
Данные ВЦИОМ (от числа респондентов, ответивших на вопрос и не выбравших вариант «Затрудняюсь ответить»), % (n = 1600)									
1. Попробую решить проблему самостоятельно	23,0	19,8	23,6	27,2	21,1	16,1	26,6	26,2	16,7
2. Попробую решить проблему вместе с друзьями, родственниками, знакомыми	24,6	25,8	38,1	18,5	38,0	12,3	18,2	20,2	8,3
3. Обращусь за помощью к профессионалам	52,3	54,4	38,3	54,4	40,9	71,7	55,1	53,6	75,0
Данные «Одноклассники.ру», % (n = 4808)									
1. Попробую решить проблему самостоятельно	19,6	13,9	24,1	32,1	14,3	19,8	18,6	23,4	30,5
2. Попробую решить проблему вместе с друзьями, родственниками, знакомыми	14,9	17,9	11,6	9,8	18,9	12,3	17,7	18,1	13,7
3. Обращусь за помощью к профессионалам	65,5	68,1	64,2	58,1	66,8	67,9	63,6	58,5	55,8

Таблица 2
**РАНЖИРОВАНИЕ ФЕДЕРАЛЬНЫХ ОКРУГОВ ПО УМЕНЬШЕНИЮ ЗНАЧЕНИЙ ПО ТРЕМ
 ВИДАМ ПОВЕДЕНИЯ ДЕЯТЕЛЬНОСТИ ЛЮДЕЙ В СЛУЧАЕ ВОЗНИКНОВЕНИЯ УГРОЗЫ ДАННЫМ
 НА ИХ ЭЛЕКТРОННОМ УСТРОЙСТВЕ**

Показатель	Источник	ЦФО	СЗФО	ЮФО	СКФО	ПФО	УФО	СФО	ДФО	КФО
1	ВЦИОМ	5	7	4	1	6	9	2	3	8
	ОК	6	9	3	1	8	5	7	4	2
2	ВЦИОМ	4	3	1	6	2	8	7	5	9
	ОК	5	3	8	9	1	7	4	2	6
3	ВЦИОМ	7	4/5	9	4/5	8	2	3	6	1
	ОК	4	1	5	8	3	2	6	7	9

дальнейших усилий по совершенствованию инструментария, а также учета всех существующих ограничений, они имеют все шансы занять свою нишу в арсенале методов социальных и маркетинговых исследований.

ЛИТЕРАТУРА

1. Шариков А.В. О четырехкомпонентной модели цифровой грамотности // Журнал исследований социальной политики. 2016. № 1. С. 87–98.
2. Всероссийское исследование «Индекс цифровой грамотности» [Электронный ресурс]. URL: http://www.mindex.rocit.ru/ICG_bright_color_print.pdf (дата обращения: 01.04.2016)
3. Gilster P. Digital Literacy. N.Y.: Wiley, 1997.
4. Буххорст А. Медиа- и информационная грамотность и ее «подруги» // Медиа- и информационная грамотность в обществах знания / Под ред. Е.И. Кузьмина, А.В. Паршаковой. М.: МЦБС, 2013. С. 35–44.
5. Alcock M., Fisher M.L., Hargadon S. Mastering Digital Literacy. Bloomington, IN: Solution Tree Press, 2014.
6. Critical Digital Literacies as Social Praxis: Intersections and Challenges / Ed. J. Avila, J. Z. Pandya. N.Y.: Peter Lang, 2013.
7. Banzato M. Digital literacy: cultura ed educazione per la società della conoscenza. Milan: B. Mondadori, 2011.
8. Current Trends and Future Practices for Digital Literacy and Competence / Ed. A. Cartelli. Hershey, PA: Information Science Reference, 2012.
9. MediaSmarts [Электронный ресурс]. URL: <http://mediasmarts.ca> (дата обращения: 01.04.2016)
10. Report of the Digital Britain Media Literacy Working Group. London: Ofcom, 2009.
11. Martin A. DigEuLit – a European Framework for Digital Literacy: a Progress Report // Journal of eLiteracy. 2005. Vol. 2. P. 130–136.
12. Jenkins H., Purushotma R., Weigel M. et al. Confronting the Challenges of Participatory Culture: Media Education for the 21st Century // Jenkins H., Purushotma R., Weigel M., Klinton K., Robinson A. The John D. and Catherine T MacArthur Foundation Reports on Digital Media and Learning. Cambridge, MA; London: The MIT Press, 2009.
13. Fundacja Nowoczesna Polska [Электронный ресурс]. URL: <https://nowoczesnapolska.org.pl> (дата обращения: 01.04.2016).
14. Цифровое будущее. Каталог навыков медиа- и информационной грамотности. М.: Межрегиональный центр библиотечного сотрудничества (МЦБС), 2013.

15. Солдатова Г.У., Рассказова Е.И. Психологические модели цифровой компетентности российских подростков и родителей // Национальный психологический журнал. 2014. № 2(14). С. 27–35.

16. Ilomaki L., Lakkala M., Kantosalo A. What is Digital Competence? // Linked Portal. Brussels: European Schoolnet (EUN), 2011. P. 1–12.

17. Описание построения выборочной совокупности [Электронный ресурс]. URL: <http://wciom.ru/research/research/omnibus/vyboroka/> (дата обращения: 01.04.2016).

18. Джанек Л., Крейтер Ф., Берг М. и др. Отчет AAPOR о больших данных / Американская ассоциация исследователей общественного мнения; пер. с англ. Д. Рогозина, А. Ипатовой, Е. Вьюговской. М., 2015.

19. Мавлетова А.М. Социологические опросы в сети Интернет: возможности построения типологии // Социология: методология, методы, математическое моделирование. 2010. № 31. С. 115–134.

Davydov Sergey

National Research University Higher School of Economics (NRU HSE),
Moscow, sdavydov@hse.ru

Logunova Olga

National Research University Higher School of Economics (NRU HSE),
Moscow, ologunova@hse.ru

“Digital literacy index” project: Methodological experiments

Within the “Digital literacy index” project we conducted a series of methodological experiments based on the analysis of «big data» and data obtained from a survey of social networks users. We found significant differences between the data obtained in the WCIOM survey and the experimental results. The analysis of «big data» showed a more even distribution of search activity and users’ interest in online media by federal districts. Depending on the type of data source the different “top” and “outsider” regions were found. The survey of the social network users also showed significant differences in the examined indicators in comparison with the offline WCIOM opinion poll – up to the diametrically opposite estimatets at the level of federal districts. The article discusses the possible mechanisms causing these differences.

Key words: digital literacy, big data, mass survey

References

1. Sharikov A.V. O chetyrehkomponentnoj modeli cifrovoj gramotnosti (in Russian), *Zhurnal issledovaniy social'noj politiki*. 2016. № 1. S. 87–98.
2. *Vserossijskoe issledovanie «Indeks cifrovoj gramotnosti»* (in Russian). URL: http://www.mindex.rocit.ru/ICG_bright_color_print.pdf (date of access: 01.04.2016)
3. Gilster P. *Digital Literacy*. N.Y.: Wiley, 1997.
4. Bukhorst A. Media- i informacionnaya gramotnost' i ee «podrugi» (in Russian), in: *Media- i informacionnaya gramotnost' v obshhestvah znaniya*. M.: MCBS, 2013. S. 35–44.
5. Alcock M., Fisher M.L., Hargadon S. *Mastering Digital Literacy*. Bloomington, IN: Solution Tree Press, 2014.
6. Avila J., Pandya J.Z. (eds.) *Critical Digital Literacies as Social Praxis: Intersections and Challenges*. N.Y.: Peter Lang, 2013.

7. Banzato M. *Digital literacy: cultura ed educazione per la società della conoscenza*. Milan: B. Mondadori, 2011.
8. Cartelli A. *Current Trends and Future Practices for Digital Literacy and Competence*, Hershey, PA: Information Science Reference, 2012.
9. *MediaSmarts*. URL: <http://mediasmarts.ca> (date of access: 01.04.2016)
10. *Report of the Digital Britain Media Literacy Working Group*. London: Ofcom, 2009.
11. Martin A. DigEuLit – a European Framework for Digital Literacy: a Progress Report, *Journal of eLiteracy*, 2005, 2, 130–136.
12. Jenkins H., Purushotma R., Weigel M. et al. Confronting the Challenges of Participatory Culture: Media Education for the 21st Century, in: Jenkins H., Purushotma R., Weigel M., Klinton K., Robinson A. *The John D. and Catherine T MacArthur Foundation Reports on Digital Media and Learning*. Cambridge, Massachusetts; London, England: The MIT Press, 2009.
13. *Fundacja Nowoczesna Polska*. URL: <https://nowoczesnapolska.org.pl> (date of access: 01.04.2016).
14. *Cifrovoe budushhee. Katalog navykov media- i informacionnoj gramotnosti* (in Russian). M.: Mezhhregional'nyj centr bibliotechnogo sotrudnichestva (MCBS), 2013.
15. Soldatova G.U., Rasskazova E.I. Psihologicheskie modeli cifrovoj kompetentnosti rossijskih podrostkov i roditel'ej (in Russian), *Nacional'nyj psihologicheskij zhurnal*. 2014. №2 (14). S. 27–35.
16. Iilomaki L., Lakkala M., Kantosalo A. What is digital competence? *Linked portal. Brussels: European Schoolnet (EUN)*, 2011, 1–12.
17. *Opisanie postroeniya vyborochnoj sovokupnosti* (in Russian). URL: <http://wciom.ru/research/research/omnibus/vyborka/> (date of access: 01.04.2016).
18. Dzhapek L., Krejter F., Berg M. i dr. *Otchet AAPOR o bol'shikh dannyh / Amerikanskaya associaciya issledovatelej obshhestvennogo mneniya*. M., 2015.
19. Mavletova A.M. Sociologicheskie oprosy v seti Internet: vozmozhnosti postroeniya tipologii (in Russian), *Sotsiologiya 4M (Sociology: methodology, methods, mathematical modeling)*, 2010, 31, 115–134.