

УДК 316.1:004

Оценка эффективности применения цифровых технологий для сбора и обработки социологических данных: критериальный подход**Толкачева Е.В.**

Кандидат социологических наук, доцент,
доцент кафедры социального анализа
и математических методов в социологии
Санкт-Петербургского государственного университета

Распространение цифровых технологий способствует развитию научных знаний в различных областях, связанных со сбором данных, включая социологию. В то же время наблюдаются противоречия, заключающиеся в необходимости ускорения процессов сбора и анализа социальных данных, с одной стороны, при обеспечении высокого качества получаемых данных, а также защиты от потенциальных рисков и угроз, связанных с использованием цифровых технологий, с другой стороны. Данная работа направлена на углубление понимания того, как цифровые технологии могут быть включены в традиционные методы опроса, анализа документов и наблюдения. В результате проведенного исследования были определены критерии оценки эффективности цифровых технологий, используемых при сборе и анализе социологических данных. Выделены основные этапы и процедуры методики оценки эффективности.

Ключевые слова: эффективность, цифровые технологии, социологическое исследование, кибербезопасность, сбор и обработка данных

Применение цифровых технологий способствует совершенствованию процессов автоматизированного сбора и обработки социологических данных. Автоматизация позволяет не только освободить социологов от непосредственного участия в рутинном процессе обработки данных, но может рассматриваться также как способ сокращения времени и материальных ресурсов, затрачиваемых для сбора данных в прикладной социологии. Появление большого количества цифровых технологий и прикладного программного обеспечения, которые могут быть использованы социологами для сбора и обработки данных, создает ситуацию некоторой «растерянности» перед обилием новых методов в процессе принятия решения о выборе средств познавательной деятельности [1].

Важным является тот факт, что сбор и обработка социологических данных с помощью цифровых технологий должны учитывать существующие требования к качеству данных. В данном случае значимым является то, насколько используемый цифро-

вой инструмент позволяет получить устойчивые и достоверные данные, способствует сокращению количества пропусков и искажений в них, в том числе путем соблюдения требований кибербезопасности.

Цель данного исследования – определить критерии оценки эффективности цифровых технологий, используемых при сборе и анализе социологической информации. В качестве теоретико-методологической основы исследования выступили работы отечественных и зарубежных авторов в области цифровых технологий и методологии социологического исследования. В ходе исследования были использованы методы сравнительного анализа, обобщения, формализации, конкретизации.

Определимся с используемыми в статье терминами и их трактовками. Цифровые технологии характеризуются нами по способу передачи данных в виде последовательности битов – нулей и единиц. При этом цифровые технологии включают в себя не только цифровые устройства, способные принимать и обрабатывать цифровой сигнал, но и приклад-

ные компьютерные программы, которые содержат основные алгоритмы операций с данными. Соответственно, тогда, когда речь идет о применении цифровых технологий в процессах сбора и обработки социологических данных, в первую очередь ставится вопрос, связанный с выбором цифровых устройств и прикладных программ, решающих поставленные исследовательские задачи на требуемом уровне качества и с меньшими затратами. То есть встает вопрос выбора наиболее эффективного способа организационно-технологического обеспечения процесса сбора и обработки социологических данных с помощью прикладных методов социологического исследования, таких как опрос, анализ документов, наблюдение.

Виды цифровых технологий, используемых в процессе сбора и обработки социологических данных, можно отнести к группе технологий так называемого «человеко-машинного взаимодействия» [2]. Это – системы распознавания образов, распознавания и синтеза речи, автоматического транскрибирования аудио. Составляющими компонентами данных систем являются устройства сбора данных: камеры, микрофоны, сканеры, датчики.

В данной статье рассматривается эффективность цифровых технологий, которые могут быть использованы в ходе наблюдения, экспертного опроса, анализа документов, интервьюирования, анкетирования. Эффективность цифровых технологий в контексте данной статьи трактуется как характеристика, обозначающая свойство цифровых технологий, используемых для сбора и обработки социологических данных, предполагающее соотношение между достигнутыми результатами и затраченными временными, материальными и трудовыми ресурсами. Обозначает способность достижения целей социологического исследования с заданным качественным уровнем, которое обеспечивается в том числе путем соблюдения требований кибербезопасности.

Каждый из социологических методов сбора данных предполагает, как правило, интеграцию нескольких цифровых технологий. Так, метод наблюдения может быть реализован с помощью интеграции систем распознавания образов, распознавания речи и транскрибирования аудио – с использованием микрофонов, камер, датчиков. Применение метода экспертного опроса предполагает использование систем распознавания и синтеза речи, транскрибирования аудио, а также микрофонов и аудиоколонок. Метод анализа документов – систем распознавания образов и сканера.

Проведем сравнительный анализ эффективности цифровых технологий в процессах сбора и обработки социологических данных. Эффективность цифровых технологий предполагает соответствие собранных данных требованиям достоверности и полноты, которые обеспечиваются в том числе пу-

тем предотвращения нарушений технологий обработки данных, отказов и сбоев цифровых устройств, ошибок в работе используемых прикладных программ.

Для количественной оценки эффективности применения цифровых технологий необходимо обозначить критерии, которые позволяют оценить достигнутые результаты и понесенные затраты. Основой для формирования критериев выступают требования, предъявляемые к качеству собранных данных. Методика оценки эффективности цифровых технологий основана на сравнительном анализе экономичности применяемых технологий. Порядок оценки эффективности включает в себя три этапа.

Первый этап предполагает проведение пилотажного исследования с использованием двух процедур сбора и обработки данных: с применением цифровых технологий и без них. В ходе данного этапа фиксируются основные временные, материальные и трудовые затраты по каждой процедуре сбора и обработки социологических данных в отдельности.

Второй этап включает в себя оценку качества собранных и обработанных данных с последующим кодированием полученных результатов. Полученные результаты характеризуются по критериям полноты и достоверности следующим образом:

код 01 присваивается, если данные не распознаны, система выдала ошибку распознавания либо данные неверны более чем на 50% независимо от полноты распознавания;

код 02 – данные частично распознаны, частично с ошибками;

код 03 – данные распознаны в пределах 50 %, распознанная часть содержит до 5 % ошибок;

код 04 – данные распознаны в пределах 95 % и более, распознанная часть содержит до 5 % ошибок;

код 05 – полнота распознавания данных и количество ошибок распознавания до 5 %.

На третьем этапе проводится сравнительный анализ затрат на сбор и обработку социологических данных с применением цифровых технологий. Затраты на применение цифровых технологий при сборе данных состоят из задействованных временных, материальных и трудовых ресурсов. При оценке эффективности применения цифровых технологий затраты необходимо сопоставить с затратами, которые были понесены при использовании социологического метода без цифровых технологий. Структура временных затрат включает в себя время, необходимое для оценки эффективности применения цифровой технологии на пилотажном этапе социологического исследования, времени сбора и обработки данных, а также времени, затраченном на исправление созданных ошибок и устранение возможных сбоев в работе цифровых устройств и прикладных программ. Необходимо отметить, что в результате проведенного сравнительного анализа

могут быть сделаны выводы, что затраты на применение цифровых технологий превышают затраты при использовании традиционной процедуры сбора и обработки данных. Оптимальным же можно считать затраты до 50 % от затрат, понесенных при использовании традиционных методов. Для оценки эффективности применения той или иной цифровой технологии полученные результаты и затраты, связанные с ее использованием, сопоставляются друг с другом. Если затраты ниже 50 % относительно затрат на применение классических методов, а также если собранным и обработанным данным был присвоен код 04 или 05, то применение цифровых технологий можно считать достаточно эффективным.

Опишем возможные причины низкой эффективности цифровых технологий при сборе и обработке социологических данных. Применение систем распознавания образов предполагает риски получения недостоверных и неполных данных в случае неисправности камеры или шумов во внешней среде. Часть информации может быть искажена вследствие неверного расположения объекта исследования, плохого освещения, дефекта печати бумажных документов или ее низкого разрешения, геометрического искажения. Могут возникать сложности в распознавании сленговых фраз и неформального языка. Плохо распознаются: художественные шрифты; вертикальный текст, анкеты, где каждый символ пишется в отдельной ячейке; очень крупный текст, например, когда одно слово занимает половину изображения [3]. Также системы могут содержать ограничения по работе с файлами. Так, например, большинство систем успешно распознает файлы форматов *jpeg*, *png* и *pdf* с размером файлов не больше 20 MB и не больше 300 страниц в файле.

При использовании системы распознавания речи следует учитывать, что вид опроса может влиять на достоверность собранной информации. В случае массового опроса формулировки вопросов ориентированы на различные группы респондентов, в связи с чем в словарях системы распознавания речи может быть достаточное количество примеров для успешного распознавания. Гораздо сложнее обстоит ситуация с экспертным опросом, который предполагает применение специальной терминологии, что влияет на точность распознавания речи эксперта. Качество звука, темп и четкость произношения могут негативно сказываться на конечном результате распознавания текста [4]. Речь может содержать неоднозначные или нечеткие выражения, которые могут затруднить их распознавание. Например, одно и то же слово или фраза могут иметь разные значения в разных контекстах. У системы распознавания речи могут возникнуть трудности в определении правильного значения и интерпретации таких выражений.

Системы синтеза речи могут выдавать ошибки вследствие неверного распознавания, полу-

ченного от респондента сообщения и генерирования неверного отклика. Похожая проблема может возникнуть и в случае применения системы автоматического транскрибирования аудио.

Результаты проведенного исследования позволяют сделать ряд выводов. Применение цифровых технологий в процессах сбора и обработки социологических данных не всегда может способствовать снижению затрат времени и количества ошибок данных. Одним из этапов оценки эффективности является фиксация затрат на процессы сбора и обработки данных с применением цифровых технологий. Предложенная в статье методика оценки эффективности цифровых технологий предполагает сопоставление результатов сбора и обработки данных с результатами, полученными традиционными методами с помощью ручного сбора и ввода данных. Использование любых цифровых технологий актуализирует вопросы, касающиеся соблюдения требований кибербезопасности в части сохранения целостности и достоверности собранных и обработанных данных. Применение цифровых технологий в социологических исследованиях может быть достаточно эффективным. В то же время сохраняется множество неурегулированных вопросов этического, правового, экономического, процедурного и инфраструктурного плана, что может быть предметом дальнейших научных исследований, обсуждений и дискуссий.

Литература:

1. Татарова Г.Г. Методологическая травма социолога. К вопросу интеграции знания // Социологические исследования. – 2006. – № 9. – С. 3–12.
2. Доронин А.М., Романов Д.А., Романова М.Л. Человеко-машинное взаимодействие и его показатели // Вестник Адыгейского государственного университета. – 2005. – № 4. – С. 244–249.
3. Измалкова А.И. Задача визуального распознавания слов при чтении на родном и на иностранном языках // Вестник Московского государственного лингвистического университета. Образование и педагогические науки. – 2014. – № 16 (702). – С. 34–41.
4. Бабаринов С.Л., Будникова М.А. О распознавании речи // Экономика. Информатика. – 2014. – № 21-1 (192). – С. 182–185.

Assessing the Efficacy of Utilizing Digital Technologies for the Collection and Processing of Sociological Data: a Criterion-Based Approach

Tolkacheva E. V.
Saint Petersburg State University

The development of digital technologies contributes to the development of scientific knowledge in various fields, including sociology. But there is a contradiction, which lies in the need to accelerate the collection and processing of sociological data while maintaining the high quality of the data obtained, as well as protection from potential risks and threats associated with the use of these technologies. This research provides a deeper understanding of how digital technologies can be integrated into traditional methods of interviewing, document analysis and surveillance. As a result, criteria for evaluating the effectiveness of digital technologies used in the collection and analysis of sociological data were determined. The main stages and procedures of the effectiveness assessment methodology are highlighted.

Key words: efficiency, digital technologies, sociological research, cyber security, data collection and processing

