

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра социальной информатики

**РАЗРАБОТКА ПРОГРАММНОГО ИНСТРУМЕНТА ДЛЯ  
АВТОМАТИЗИРОВАННОГО КОДИРОВАНИЯ  
СОЦИОЛОГИЧЕСКИХ ДАННЫХ**

(автореферат бакалаврской работы)

студента 4 курса 451 группы  
направления 09.03.03 - Прикладная информатика  
профиль Прикладная информатика в социологии  
Социологического факультета  
Павлова Никиты Алексеевича

Научный руководитель  
кандидат социологических наук, доцент

\_\_\_\_\_ К.В. Мохнаткина  
подпись, дата

Зав. кафедрой  
кандидат социологических наук, доцент

\_\_\_\_\_ И.Г. Малинский  
подпись, дата

Саратов 2026

## ВВЕДЕНИЕ

**Актуальность проблемы.** В условиях цифровизации научных исследований и широкого распространения различных онлайн-инструментов для сбора данных особую актуальность приобретает задача эффективной обработки результатов социологических опросов. Одним из самых популярных способов получения эмпирических данных является использование веб-сервисов анкетирования, таких как Google Forms, Yandex Forms и их аналогов. Они обеспечивают быстрый сбор ответов респондентов. Сервисы позволяют совершить экспорт данных в табличном виде.

Полученные данные, в большинстве случаев, представлены в текстовом формате и требуют предварительной обработки перед проведением статистического анализа. Одним из ключевых шагов является кодирование данных. Оно подразумевает присвоение текстовым ответам соответствующих числовых значений, пригодных для дальнейшего анализа в специализированных программных пакетах, например, в SPSS.

На практике кодирование данных осуществляется вручную, либо же с помощью функции замены в табличных редакторах. Методы имеют весомые недостатки: высокая трудоёмкость, большие временные затраты, риск допущения ошибок. При масштабировании исследования они становятся критическими.

Автор столкнулся с проблемами выше при обработке результатов анкетирования. Ручное кодирование оказалось неэффективным, что вызвало необходимость поиска более оптимального решения.

Актуальность выпускной квалификационной работы обусловлена необходимостью автоматизации процесса кодирования социологических данных, с целью повышения эффективности их обработки, сокращения временных затрат и минимизации ошибок. В рамках работы разрабатывается авторский программный инструмент «СоциоКодер», ориентированный на практическое применение в обработке данных.

**Степень научной разработанности данной проблемы.** Связана с развитием эмпирической социологии и количественного анализа. Базовые принципы были сформированы в рамках контент-анализа и прикладной социологии, затем они получили дальнейшее развитие в подходах к обработке качественных данных и цифровых технологиях анализа информации. Тем не менее, задачи автоматизации кодирования сохраняют свою актуальность и требуют совершенствования, несмотря на накопленный теоретический и практический опыт.

**Цель** исследования заключается в разработке авторского программного инструмента для автоматизированного кодирования социологических данных.

Согласно цели данной работы необходимо решить следующие **задачи**:

1. Изучить и представить методологические основы автоматизированного кодирования социологических данных;
2. Описать теоретические основания проектирования программного инструмента для кодирования социологических данных;
3. Сформировать список требований для авторского программного продукта автоматизированного кодирования социологических данных;
4. Создать и реализовать программный продукт «СоциоКодер».
5. Оценить эффективность применения «СоциоКодер» в изучении качества технического оснащения учебных аудиторий социологического факультета по оценкам студентов.

**Объект** данной работы – программный продукт «СоциоКодер» для автоматизации кодировки социологических данных.

**Предмет** данной работы – эффективность авторской программы «СоциоКодер» в изучении качества технического оснащения аудиторий социологического факультета.

**Практическая значимость исследования** заключается в разработке авторского программного инструмента, обеспечивающего автоматизацию процесса кодирования социологических данных.

**Структура выпускной квалификационной работы** представлена введением, тремя разделами, заключением, списком использованных источников, приложениями А и Б.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**В первом разделе «Методологические основы автоматизированного кодирования социологических данных»** рассматриваются степень изученности проблемы кодирования, особенности данных, получаемых с помощью онлайн-сервисов анкетирования, типы вопросов и шкалы измерения, их влияние на процесс кодирования, теоретические подходы к кодированию и методология автоматизированного кодирования.

Проблема кодирования данных была сформирована в рамках развития эмпирической социологии и связана с необходимостью преобразования качественных смыслов в количественные единицы для дальнейшего подсчёта. Основы количественного подхода были заложены Г. Лассуэллом<sup>1</sup>, Б. Берельсоном<sup>2</sup> и П. Лазарсфельдом<sup>3</sup>.

На этапе становления социологии данные обрабатывались преимущественно руками. Подобный способ кодирования используется и по сей день. В рамках «обоснованной теории» Б. Глейзера и А. Страусса<sup>4</sup> кодирование предполагает выделение смысловых единиц и их дальнейшую категоризацию. Вопросы изменения процессов сбора и обработки

---

<sup>1</sup> Лассуэлл, Г. Д. The language of politics: studies in quantitative semantics / Harold D. Lasswell, Nathan Leites and associates. — New York : George W. Stewart, 1949. — 271 p.

<sup>2</sup> Berelson, B. Content analysis in communication research / By Bernard Berelson .... — [New York] : Free press of Glencoe, [1952]. — 220 с. диагр.; 22 см. — (Foundations of communication research).

<sup>3</sup> Лазарсфельд, П. Ф. Qualitative analysis: historical and critical essays / Paul F. Lazarsfeld. — Boston : Allyn and Bacon, 1972. — xvii, 457 p.

<sup>4</sup> Glaser, B. G. The Discovery of Grounded Theory: Strategies for Qualitative Research / B. G. Glaser, A. L. Strauss. — Chicago: Aldine Publishing Company, 1967. — 271 p.

социологической информации освещены в работах Е.В. Толкачёвой<sup>5</sup> и Г.В. Антонова<sup>6</sup>.

Стремительное развитие цифровых технологий и массовое использование онлайн-сервисов (Google Forms, Yandex Forms) привели к существенным изменениям в процессах сбора информации и увеличению объёмов данных. В современных реалиях происходит переход к частичной автоматизации: социологи применяют системы класса CAQDAS, методы обработки естественного языка (NLP) и большие языковые модели (LLM). Однако автоматические алгоритмы могут допускать ошибки, что сохраняет за исследователем необходимость контролировать процесс кодирования.

Особое внимание уделяется специфике социологических данных, полученных онлайн. Их отличительной чертой является неоднородность. Возникает проблема пропущенных значений, которым для корректной работы статистических пакетов присваиваются специальные служебные коды. Классификация вопросов (открытые и закрытые) и выбор шкалы измерения (номинальная, порядковая, количественная) напрямую определяют характер дальнейшей обработки и границы допустимой интерпретации социологической информации.

Представлена методология автоматизированного кодирования, направленная на систематизацию, ускорение и повышение точности обработки данных. Она включает в себя следующие этапы: выделение уникальных вариантов ответов, подсчёт частоты их встречаемости, интерактивное

---

<sup>5</sup> Толкачева, Е. В. Особенности процессов сбора и обработки социологических данных с применением цифровых технологий // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 11. Социология. — 2024. — № 1 [Электронный ресурс]. — URL: <https://cyberleninka.ru/article/n/osobennosti-protsessov-sbora-i-obrabotki-sotsiologicheskikh-dannyh-s-primeneniem-tsifrovyyh-tehnologiy>

(дата обращения: 03.04.2026). — Загл. с экрана. — Яз. рус.

<sup>6</sup> Антонов, Г. В. Особенности компьютерной обработки социологических данных // Logos et Praxis. — 2006. — № 5 [Электронный ресурс]. — URL: <https://cyberleninka.ru/article/n/osobennosti-kompyuternoy-obrabotki-sotsiologicheskikh-dannyh>.

(дата обращения: 03.04.2026). — Загл. с экрана. — Яз. рус.

назначение пользователем числовых кодов, обработка пропусков и пользовательская интерпретация характера отсутствия ответа.

В первом разделе была рассмотрена степень изученности проблемы кодирования социологических данных и теоретико-методологические основы их обработки. Выявлено, что эволюция эмпирических методов исследования, становление контент-анализа и внедрение цифровых технологий привели к развитию подходов к кодированию. Это стало причиной перехода от ручных методов обработки информации к частичной автоматизации.

Также, проанализированы особенности социологических данных, типы вопросов и шкалы измерения. Доказано, что кодирование является важным этапом подготовки данных к статистическому анализу. Существующие методы имеют ряд недостатков, например, трудоёмкость и риск ошибок. В результате была теоретически обоснована необходимость разработки авторского программного инструмента для автоматизированного кодирования социологических данных, сочетающего обработку данных алгоритмом с интерактивным участием пользователя.

**Во втором разделе «Теоретические основания проектирования программного инструмента для кодирования социологических данных»** формируются требования к программному инструменту, его архитектура, алгоритм работы, модель данных и особенности пользовательского интерфейса.

На основе методологической базы был сформирован перечень функциональных требований. Ключевыми среди них являются: загрузка исходных данных в формате CSV, отображение структуры анкеты, выявление уникальных вариантов ответов и их частоты, интерактивное присвоение числовых кодов, контроль полноты кодирования, сохранение результатов, формирование и хранение словаря кодирования, автоматическое предложение готовых шкал, обработка пропущенных значений, экспорт закодированных данных.

Определены входные и выходные данные, а также ограничения системы. В частности, обоснован осознанный отказ от использования методов

интеллектуального анализа текста (NLP) для автоматической группировки открытых ответов, это сделано для обеспечения точности интерпретации данных и сохранения контроля над процессом со стороны исследователя.

Спроектирована модульная, событийно-ориентированная архитектура приложения, включающая восемь базовых модулей. Описан пошаговый алгоритм автоматизированного кодирования, который наглядно демонстрирует, как текстовый массив преобразуется в числовой формат, пригодный для анализа в статистических пакетах (например, SPSS).

Представлена модель хранения данных на основе таблиц DataFrame и словаря кодирования в формате JSON, обеспечивающего повторное использование сохранённых шкал кодирования. Спроектирован интуитивно понятный графический интерфейс. Для повышения продуктивности работы социолога и снижения риска случайных ошибок в интерфейсе заложены интерактивное редактирование данных, автоматическая фокусировка ввода, поддержка клавиатурной навигации, контроль полноты ввода данных, система уведомлений, использование ранее сохранённых шкал.

Во втором разделе был проведён анализ требований к программному инструменту для автоматизированного кодирования социологических данных. Определены функциональные возможности системы, от загрузки файла до экспорта закодированных данных. Помимо этого, были установлены ограничения, которые связаны со структурой входных данных и необходимостью участия пользователя в процессе назначения кодовых значений.

На основе выявленных требований была разработана архитектура программного инструмента. Она основана на модульном и событийно-ориентированном подходах. Описаны основные программные модули и их взаимодействие через внутренние структуры данных, таких как DataFrame и словарь кодирования. Кроме того, был представлен алгоритм автоматизированного кодирования социологических данных. Он отражает

последовательность этапов обработки от загрузки исходного файла до экспорта массива кодовых значений.

Также, были определены модель данных и принципы хранения кодировочных схем, обеспечивающих возможность их повторного использования. На этапе проектирования пользовательского интерфейса была сформирована логичная и интуитивно понятная структура взаимодействия пользователя с системой.

**В третьем разделе «Апробация и оценка эффективности программного инструмента кодирования социологических данных на примере изучения качества технического оснащения учебных аудиторий социологического факультета по оценкам студентов»** описываются создание и реализация программного продукта «СоциоКодер», а также проводится оценка эффективности его применения.

Реализация программного инструмента осуществлялась на языке программирования Python. Выбор языка обусловлен простотой его синтаксиса, наличием всех необходимых библиотек для работы с данными и интерфейсом, а также широким применением в аналитических задачах.

Работа программы включает загрузку CSV-файла с результатами анкетирования, автоматическое формирование структуры анкеты, выделение уникальных вариантов ответов и расчёт частоты их встречаемости. Пользователь присваивает ответам числовые коды, после чего программа автоматически применяет их ко всему массиву данных и формирует итоговый набор закодированных данных.

В ходе тестирования была подтверждена стабильность работы программного инструмента и его соответствие заявленным требованиям. Проведённый сравнительный анализ программного и ручного кодирования показал, что использование продукта «СоциоКодер» позволяет сократить время обработки данных более чем в 3 раза: кодирование массива данных заняло 1 минуту 20 секунд против 4 минут 20 секунд при ручной обработке в Excel. Также, не было выявлено ошибок кодирования благодаря встроенным

механизмам контроля корректности ввода и автоматическому применению кодов. Кроме того, программный инструмент обеспечивает более удобную и последовательную обработку данных по сравнению с ручным кодированием.

Апробация программного инструмента проводилась на материалах исследования качества технического оснащения учебных аудиторий социологического факультета по оценкам студентов. Полученные результаты показали, что использование программы «СоциоКодер» позволяет значительно ускорить процесс кодирования социологических данных: среднее время обработки массива составило 1 минуту 20 секунд против 4 минут 20 секунд при ручном кодировании. Это подтверждает эффективность разработанного программного инструмента и целесообразность его применения в исследовательской практике.

В рамках работы над третьим разделом был реализован авторский программный инструмент автоматизированного кодирования социологических данных на языке Python<sup>7</sup> с использованием библиотек Tkinter<sup>8</sup>, pandas<sup>9</sup>, os<sup>10</sup>, json<sup>11</sup>. Разработанная программа обеспечивает загрузку данных в формате CSV, выделение уникальных ответов с подсчётом частоты их встречаемости, присвоение числовых кодов в интерактивной форме, а также сохранение кодовых схем в словарь. Пользовательский интерфейс даёт возможность обрабатывать вопросы анкеты последовательно и контролировать полноту кодирования, это делает процесс работы с данными более удобным.

---

<sup>7</sup> Саммерфилд, М. Программирование на Python 3 : подробное руководство / пер. с англ. — СПб. : Символ-Плюс, 2009. — 608 с. : ил.

<sup>8</sup> tkinter — Python interface to Tcl/Tk — Python 3.14.5rc1 documentation [Электронный ресурс]. — URL: <https://docs.python.org/3/library/tkinter.html> (дата обращения: 13.04.2026). — Загл. с экрана. — Яз. англ.

<sup>9</sup> User Guide — pandas 3.0.2 documentation [Электронный ресурс]. — URL: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/index.html#user-guide](https://pandas.pydata.org/pandas-docs/stable/user_guide/index.html#user-guide) (дата обращения: 13.04.2026). — Загл. с экрана. — Яз. англ.

<sup>10</sup> json — JSON encoder and decoder — Python 3.14.5rc1 documentation [Электронный ресурс]. — URL: <https://docs.python.org/3/library/json.html> (дата обращения: 13.04.2026). — Загл. с экрана. — Яз. англ.

<sup>11</sup> os — Miscellaneous operating system interfaces — Python 3.14.5rc1 documentation [Электронный ресурс]. — URL: <https://docs.python.org/3/library/os.html> (дата обращения: 13.04.2026). — Загл. с экрана. — Яз. англ.

По результатам тестирования было выявлено, что программный инструмент обрабатывает входные данные корректным образом, устойчиво работает при наличии пропусков и обеспечивает сохранение результатов кодирования без потерь информации. Благодаря механизму повторного применения ранее заданных шкал процесс обработки данных дополнительно ускоряется. Функциональные возможности системы полностью соответствуют установленным требованиям и обеспечивают выполнение основных операций для подготовки данных к дальнейшему статистическому анализу.

Также, сравнительный анализ подтвердил преимущество разработанного программного инструмента перед традиционным ручным кодированием. Это связано с тем, что автоматизация этапов работы позволила сократить время обработки данных более чем в 3 раза и снизить риск ошибок, связанных с человеческим фактором. Отсюда можно сделать вывод о практической эффективности авторского решения и целесообразности его использования в учебной и исследовательской деятельности.

## **ЗАКЛЮЧЕНИЕ**

В ходе выполнения выпускной квалификационной работы была рассмотрена проблема кодирования социологических данных, как важного этапа подготовки эмпирической информации к дальнейшему анализу. Проведённый анализ степени разработанности данной проблемы показал, что невзирая на наличие теоретических подходов и методических рекомендаций, процесс кодирования остаётся довольно трудоёмким и подверженным влиянию человеческого фактора. Были проанализированы особенности социологических данных, получаемых путём использования онлайн-сервисов анкетирования. Также, изучены типы вопросов анкет и их влияние на процесс кодирования. Наряду с этим, рассмотрены шкалы измерения в социологических исследованиях. Это позволило определить их роль в формировании корректных кодировочных схем. Анализ теоретических подходов к кодированию позволил обосновать необходимость автоматизации процесса и сохранении, при этом, участия исследователя.

На основе выполненного анализа была разработана методология автоматизированного кодирования социологических данных, которая сочетает в себе алгоритмическую обработку информации и участие пользователя. В рамках проектирования программного инструмента были сформулированы требования к системе и определена её архитектура, которая основана на модульном и событийно-ориентированном подходе. Также, создан алгоритм автоматизированного кодирования. Особое внимание было модели данных и хранении словаря кодирования, который обеспечивает возможность повторного применения ранее сохранённых шкал. Вместе с тем, был спроектирован пользовательский интерфейс. Он ориентирован на последовательную обработку вопросов анкеты и обеспечивает удобство взаимодействия с системой.

В рамках практической части работы был реализован авторский программный инструмент для кодирования социологических данных. Разработанная программа «СоциоКодер» выполняет такие функции, как загрузка файлов в формате CSV, выделение уникальных ответов с подсчётом частоты их встречаемости, интерактивное присвоение числовых кодов и формирование итогового набора данных. Корректность работы, устойчивость к разным типам входных данных и удобство использования программного продукта были подтверждены тестированием. Результаты проведённого сравнительного анализа показывают, что применение разработанного инструмента позволяет в значительной мере сократить время на процесс кодирования и минимизировать риск возникновения ошибок.

Все задачи, поставленные в рамках выпускной квалификационной работы, были решены в полном объёме. Цель разработки программного инструмента для автоматизированного кодирования социологических данных достигнута. Полученные результаты подтверждают практическую значимость предложенного решения и целесообразность его использования в учебной и научно-исследовательской деятельности. Перспективы дальнейшего развития

закладаются в улучшение реализованных механизмов и расширении функциональных возможностей.