

DOI: [10.14515/monitoring.2025.2.2575](https://doi.org/10.14515/monitoring.2025.2.2575)**С. В. Жучкова, В. Ю. Девятников, А. П. Казун, М. Д. Белов, О. И. Сидорова****ТЕКСТЫ СУДЕБНЫХ ПРИГОВОРОВ КАК ИСТОЧНИК ДАННЫХ
ДЛЯ ЭМПИРИЧЕСКИХ ИССЛЕДОВАНИЙ ПРАВА В РОССИИ****Правильная ссылка на статью:**

Жучкова С. В., Девятников В. Ю., Казун А. П., Белов М. Д., Сидорова О. И. Тексты судебных приговоров как источник данных для эмпирических исследований права в России // Мониторинг общественного мнения: экономические и социальные перемены. 2025. № 2. С. 170—192. <https://doi.org/10.14515/monitoring.2025.2.2575>.

For citation:

Zhuchkova S. V., Devyatnikov V. Y., Kazun A. P., Belov M. D., Sidorova O. I. (2025) Texts of Court Verdicts as a Data Source for Empirical Legal Studies in Russia. *Monitoring of Public Opinion: Economic and Social Changes*. No. 2. P. 170–192. <https://doi.org/10.14515/monitoring.2025.2.2575>. (In Russ.)

Получено: 24.02.2024. Принято к публикации: 19.02.2025.

ТЕКСТЫ СУДЕБНЫХ ПРИГОВОРОВ КАК ИСТОЧНИК ДАННЫХ ДЛЯ ЭМПИРИЧЕСКИХ ИССЛЕДОВАНИЙ ПРАВА В РОССИИ

TEXTS OF COURT VERDICTS AS A DATA SOURCE FOR EMPIRICAL LEGAL STUDIES IN RUSSIA

ЖУЧКОВА Светлана Васильевна — кандидат наук об образовании, научный сотрудник, Международный центр изучения институтов и развития, научный сотрудник, Институт образования, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

E-MAIL: szhuchkova@hse.ru

<https://orcid.org/0000-0002-4425-725X>

Svetlana V. ZHUCHKOVA¹ — PhD in Education, Research Fellow, International Center for the Study of Institutions and Development; Research Fellow, Institute of Education

E-MAIL: szhuchkova@hse.ru

<https://orcid.org/0000-0002-4425-725X>

ДЕВЯТНИКОВ Вадим Юрьевич — стажер-исследователь, Международный центр изучения институтов и развития, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

E-MAIL: vdevyatnikov@hse.ru

<https://orcid.org/0009-0009-5278-7232>

Vadim Yu. DEVYATNIKOV¹ — Research Assistant, International Center for the Study of Institutions and Development

E-MAIL: vdevyatnikov@hse.ru

<https://orcid.org/0009-0009-5278-7232>

КАЗУН Антон Павлович — кандидат социальных наук, директор, Институт анализа предприятий и рынков, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

E-MAIL: akazun@hse.ru

<https://orcid.org/0000-0002-0091-5388>

Anton P. KAZUN¹ — Cand. Sci. (Soc.), Director, Institute for Industrial and Market Studies

E-MAIL: akazun@hse.ru

<https://orcid.org/0000-0002-0091-5388>

БЕЛОВ Михаил Дмитриевич — PhD-студент, Калифорнийский Университет в Лос-Анджелесе (UCLA), Лос-Анджелес, США

E-MAIL: mikaelbelov@gmail.com

<https://orcid.org/0000-0001-8848-2726>

Mikhail D. BELOV² — PhD student

E-MAIL: mikaelbelov@gmail.com

<https://orcid.org/0000-0001-8848-2726>

СИДОРОВА Ольга Игоревна — стажер-исследователь, Международный центр изучения институтов и развития, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия

E-MAIL: oi.sidorova@hse.ru

<https://orcid.org/0009-0003-1733-1471>

Olga I. SIDOROVA¹ — Research Assistant, International Center for the Study of Institutions and Development

E-MAIL: oi.sidorova@hse.ru

<https://orcid.org/0009-0003-1733-1471>

¹ HSE University, Moscow, Russia

² University of California, Los Angeles, Los Angeles, USA

Аннотация. Развитие эмпирических исследований права в России во многом сдерживается дефицитом доступных источников данных о правоприменении. Один из таких источников, до сих пор недостаточно используемый в российских исследованиях, — публикуемые в открытом доступе тексты судебных решений, в частности судебных приговоров. Цель статьи — предложить подход к использованию текстов судебных приговоров как источника данных, который может быть востребован для проведения эмпирических исследований в сфере социологии права, социологии преступности и девиантного поведения в России.

На основе собственного опыта работы с текстами приговоров авторы описывают возможные варианты их сбора, предобработки и анализа с применением техник автоматизированного анализа текстов. В качестве иллюстрации возможностей используется недавнее исследование гендерных различий в сроках наказания за убийство по ст. 105 ч. 1 УК РФ. Обсуждение потенциала использования текстов приговоров дополняется обзором альтернативных источников судебных данных, доступных в России, а также рефлексией о преимуществах и ограничениях текстов приговоров по сравнению с другими источниками правовой информации.

Ключевые слова: веб-скрейпинг, открытые данные, правоприменение, судебная статистика, тексты приговоров, эмпирические исследования права, text mining

Благодарность. Исследование выполнено за счет гранта Российского научного фонда № 23-78-10073 «Разработка и апробация методики автоматизированного анализа текстов приговоров российских судов для социально-правовых исследований (на примере насильственных преступлений)» (см. подробнее: <https://rscf.ru/project/23-78-10073/>).

Abstract. The development of empirical legal studies in Russia is restricted by the lack of sources of disaggregated data on law enforcement available to social researchers. One of the potential sources of such data, which is still insufficiently used in Russian research, is the publicly available texts of court verdicts, in particular court sentences. This article proposes a methodology for using the texts of court verdicts as a possible data source that may be of interest to sociologists conducting empirical research on law and law enforcement in Russia.

Based on their own experience of sentencing research, the authors describe possible options for the collection, preprocessing, and analysis of such texts using text-mining techniques. As an illustration, a recent study of gender differences in sentences for murder under Article 105 Part 1 of the Criminal Code of the Russian Federation is used. The discussion of the potential of using sentencing texts is complemented by an overview of alternative sources of judicial data available in Russia, as well as a reflection on the advantages and limitations of sentencing texts compared to other sources of legal information.

Keywords: web scraping, open data, law enforcement, court statistics, court verdicts, empirical legal studies, text mining

Acknowledgments. The research was supported by a grant from the Russian Science Foundation No. 23-78-10073 “Development and approbation of the algorithm for automated analysis of the court decisions texts for socio-legal studies (based on cases of violent crimes)” (see more details: <https://rscf.ru/project/23-78-10073/>).

Введение

Эмпирические исследования права — направление, предполагающее изучение широкого круга вопросов взаимодействия права и общества с обязательной опорой на эмпирические данные [Волков, 2017]. Систематические исследования в этом направлении начались в конце XIX в. в США [Фридман, 2014] и к настоящему времени также широко представлены в Германии, Израиле, Великобритании, Австралии, Канаде¹.

В России развитие эмпирических исследований права сдерживается дефицитом доступных источников дезагрегированных данных, то есть данных, представленных на уровне отдельного юридического кейса, не сгруппированных по каким-либо основаниям. В отличие от западных стран, Россия находится в уникальной ситуации, при которой ведомства собирают большое количество информации о правоприменении, однако ученые почти не имеют к ним доступа [Титаев и др., 2018]. Из этой ситуации есть исключение — публикуемые российскими судами тексты судебных актов, в том числе судебных приговоров, которые в соответствующей академической дискуссии не раз упоминались как потенциальный источник эмпирических данных [Волков, 2017; Волков, Скугаревский, Титаев, 2016; Титаев и др., 2018]. Хотя приговоры в России публикуются уже более десяти лет, а на портале государственной автоматизированной системы «Правосудие» (далее — ГАС «Правосудие») есть доступ к нескольким миллионам приговоров, практика привлечения их как источника информации все еще редка: до сих пор есть лишь несколько примеров исследований, в которых ученые², журналисты или сотрудники НКО³ в своих выводах опираются на анализ текстов приговоров. На наш взгляд, потенциал этого источника данных существенно превышает тот круг задач, который был решен на текущий момент. Более системная работа с информацией, представленной в текстах приговоров, способна дать толчок к развитию эмпирических исследований права в России, позволив ответить на вопросы как о закономерностях совершения преступлений, так и об особенностях принятия решений судьями. В свою очередь, поиск ответов на эти вопросы важен для доказательного совершенствования уголовной политики и повышения качества работы судебной системы в России.

Цель статьи — предложить подход к использованию текстов судебных приговоров в качестве возможного источника данных для эмпирических исследований права в России. Мы хотим на основе собственного опыта 1) рассказать, как собирать тексты приговоров; 2) описать методы и техники, которые можно использовать для предобработки и анализа приговоров; 3) продемонстрировать возможность извлечения из текстов приговоров переменных, которые не доступны в агрегированной статистике; 4) обсудить ограничения, связанные с использованием этих данных. Статья охватывает только приговоры как источник данных —

¹ Перечислены топ-5 стран, кроме США, авторы из которых наиболее часто публикуются в профильном журнале — *Journal of Empirical Legal Studies* (согласно данным базы научного поиска Lens).

² См., например, [Титаев, 2014; Metsker et al., 2019; Zhuchkova, Kazun, 2023].

³ См. проект «Алгоритм света» Консорциума женских НПО, в котором на основе текстов приговоров рассчитана доля убитых в России женщин, погибших от рук партнеров или родственников. URL: <https://algorithmsveta.com/> (дата обращения: 16.12.2024).

иные источники данных о правоприменительной практике, которые также могут быть релевантны для исследователей, упоминаются в обзорной части статьи, однако не являются предметом рассмотрения в настоящей работе. Предлагаемый в статье подход может быть востребован социологами, работающими в сфере социологии права, социологии преступности и девиантного поведения, а также социально-экономического анализа правоприменения.

Статья структурирована следующим образом. Вначале представлен обзор источников данных о правоприменении, использующихся в зарубежных исследованиях и доступных в России. Затем изложены нормативные основания публикации судебных приговоров в России и имеющиеся сведения о реальных практиках публикации приговоров. На примере одного из недавних исследований описаны возможные способы сбора и предобработки приговоров, а также процесс их трансформации в массив количественных данных, пригодных для анализа. Статья завершается методологической рефлексией о преимуществах и ограничениях приговоров как источника эмпирических данных.

Обзор литературы

Ключевые источники эмпирических данных

В эмпирических исследованиях права, посвященных таким вопросам, как детерминанты преступности, традиционно есть два ключевых источника информации: официальная статистика и виктимизационные опросы.

К официальной статистике относятся различные административные записи — от обращений в полицию и государственные ведомства до статистики о судебных решениях и апелляциях. Исследователи К. Лофтин и Д. Макдауэлл выделяют шесть ключевых недостатков официальной статистики [Loftin, McDowall, 2010]. Во-первых, у разных ведомств разные алгоритмы учета, что нередко делает данные несопоставимыми между собой. Во-вторых, жители не всегда обращаются в полицию и другие органы, что оставляет большое количество неучтенных случаев. В-третьих, полиция может регистрировать не все преступления по разным причинам — от перегруженности или ненамеренного игнорирования жалоб до стремления сохранить хорошую статистику, что имеет место и в России [Paneyakh, 2014]. В-четвертых, официальные данные часто содержат пропуски или же не публикуются целиком. В-пятых, многие ведомства не регистрируют важную для ученых информацию, например национальность и образование преступника и жертвы. Наконец, искажения могут быть связаны с тем, что ведомственная статистика — политически важная информация, которая говорит об эффективности работы системы правоприменения. Исследования показывают, что некоторая информация может быть недоступной ученым из-за нежелания ведомств предоставлять ее, даже если официально такая обязанность у них есть [Addington, 2008]. Более того, факт непредоставления ведомством информации не является случайным и может быть связан с качеством правоприменения в том или ином регионе или с организационными факторами работы внутри конкретного ведомства [Nolan III et al., 2007; Paneyakh, 2014].

Виктимизационные опросы в США и Европе начали развиваться в конце 1960-х — начале 1970-х годов как реакция на недостатки официальной статисти-

стики [Rand, 2006]. Базовая идея состояла в том, что, если сравнить с официальной статистикой долю населения, сообщившего в ходе опроса об опыте столкновения с преступлением, можно узнать объем неучтенной преступности. Помимо единичных опросов, репрезентирующих отдельные города или фиксирующих отношение граждан к различным правоохранительным ведомствам, в России до недавнего времени не проводилось виктимизационных опросов [Веркеев и др., 2019]. Институт проблем правоприменения при Европейском университете в Санкт-Петербурге (далее — ИПП при ЕУ СПб) провел два опроса, соответствующих международным стандартам: в 2018 г. [Кнорре, Титаев, 2018] и в 2021 г. [Серебrenников, Титаев, 2022].

Виктимизационные опросы также не лишены недостатков. К их ограничениям относят недоучет некоторых видов преступлений [Schwartz, Vega, 2017], таких как преступления без жертв (употребление наркотиков или проституция) или тех, где жертва умирает. Другие смещения связаны с особенностями выборки социологических опросов, из-за которых некоторые категории граждан (например, бездомные или очень богатые) имеют низкие шансы стать респондентом. Виктимизационные опросы нередко игнорируют тюрьмы и другие закрытые для исследователей учреждения. Наконец, на некоторые вопросы респондентам некомфортно отвечать, особенно если опыт связан с пережитым в прошлом насилием (в таких случаях также имеет место и фактор интервьюера).

Анализ результатов, полученных с использованием официальной статистики и виктимизационных опросов в США за период с 1973 по 2008 г. [Ansari, He, 2015], показывает, что оценки экономических преступлений на основе этих двух источников постепенно сходятся. Это может свидетельствовать о совершенствовании обоих источников данных. В то же время исследователи отмечают потребность в альтернативных источниках данных [Schwartz, Vega, 2017]. К таковым можно отнести данные из СМИ и социальных сетей [Hipp et al., 2019], информацию об интернет-запросах [Ozkan, 2019] и т. п. Впрочем, эти источники данных тоже далеки от совершенства. Например, публикации о преступлениях в СМИ имеют тенденцию давать смещенные оценки в пользу тех или иных типов преступлений или социальных групп [Lundman, 2003]. Кроме того, медиа нередко выступают не столько источником информации, сколько фактором, который влияет на преступность и ее восприятие населением [Pickett et al., 2015].

Другим новым источником данных являются различные юридические тексты. С точки зрения типологии этот источник близок к официальной статистике, в том числе не свободен от некоторых ее недостатков — например, в судебном приговоре не окажется той информации, которую сотрудник суда не посчитал нужным вносить. Однако анализ юридических документов с использованием различных техник автоматизированной обработки текстов предоставляет исследователям дополнительные возможности, такие как извлечение из них новой информации, в том числе такой, которую никто намеренно не фиксировал в качестве официальной статистики [Kovalchuk et al., 2022]. Хотя в настоящем исследовании мы сфокусируемся на текстах судебных приговоров как источнике данных, анализировать с помощью этих техник можно и иные тексты, включая договоры, патенты, законодательные акты, налоговые декларации и пр.

У автоматизированного анализа текстов есть широкий спектр для применения: суммаризация текстов судебных решений [Moens, 2007], классификация [Ashley, Brüninghaus, 2009] и поиск типичных паттернов [Kaur et al., 2019], предсказание судебных решений и анализ их детерминант [Aletras et al., 2016]. С помощью подобных техник можно определить, принимают ли судьи из разных регионов одной страны при прочих равных одинаковые решения [Pina-Sánchez et al., 2019], существует ли дискриминация подзащитных или жертв по полу, расе или по национальности [Pina-Sánchez, Roberts, Sferopoulos, 2019]. Автоматизированный анализ юридических текстов — быстро развивающееся направление, в котором каждый год появляются новые подходы и алгоритмы (см., например, [Shu, Sen, 2023; Prasad, Boughanem, Dkaki, 2024]), в том числе опирающиеся на большие языковые модели.

В **Приложении 1** мы суммируем основные преимущества и недостатки трех рассмотренных нами источников данных: официальной статистики, исследовательских данных (в т. ч. виктимизационных опросов) и текстов судебных приговоров.

Международная практика публикации текстов судебных приговоров

Публикация приговоров — набирающая глобальное распространение практика, которая является частью более обширного движения о доступе к правовой информации FALM⁴ (Free Access to Law Movement). Существует множество организаций, публикующих правовую информацию. Так, на сайте Мирового института правовой информации⁵ можно найти ресурсы с юридическими текстами многих стран мира. Однако публичный доступ к правовой информации включает в себя в первую очередь сами законы и лишь потом судебные решения [Poulin, Mowbray, 2007]. Ситуация с публикацией именно судебных решений, в том числе приговоров, различается от страны к стране. Наиболее близкой к российской представляется ситуация в Китае [Ahl, Sprick, 2018] и США [Cao, Ash, Chen, 2020], где открытый доступ к текстам судебных решений регулируется соответствующими национальными законами и существуют сайты, агрегирующие миллионы судебных актов. В Нидерландах, по оценкам исследователей, лишь 5 % судебных решений находятся в открытом доступе [Hartel, van Wegberg, van Staalduinen, 2022]. В Англии тексты судебных решений не публикуются в открытом доступе, однако могут быть получены через обращение к коммерческим базам данных⁶. Вопрос о полноте и несмещенности этих данных остается открытым [Pina-Sanchez et al., 2019]. Кроме России, на постсоветском пространстве национальные базы, содержащие данные о решениях судов различных инстанций, были созданы в Казахстане [Юрченко, Сакалов, 2009] и Украине [Turkanova, 2023].

Существует несколько системных проблем с базами судебных решений и доступом к ним. Во-первых, многие базы данных фокусируются на судах высших инстанций, а решения более низких инстанций либо не присутствуют вовсе, либо собраны лишь частично. Во-вторых, подобную информацию иногда нельзя полу-

⁴ The Free Access to Law Movement — Homepage // FALM. URL: <http://falm.info/> (дата обращения: 18.02.2024).

⁵ World Legal Information Institute — Homepage // WorldLII. URL: <http://www.worldlii.org> (дата обращения: 18.02.2024).

⁶ Thelawpages — Homepage // Thelawpages. URL: <https://www.thelawpages.com/> (дата обращения: 18.02.2024).

чить бесплатно. В-третьих, многие сайты сделаны на национальных языках, что усложняет работу исследователям. В-четвертых, некоторые базы данных (например, российская ГАС «Правосудие» и казахская «ЕАИАС СО РК») недоступны из других регионов, что усложняет доступ к ним.

Источники судебных данных в России

Важный источник информации о правоприменении в России — отчеты по форме 4-ЕГС, предоставляющие агрегированные данные о состоянии преступности в регионах России. Хотя их актуальная публикация на текущий момент не осуществляется (доступны данные за 2014—2022 гг., в связи с доработкой функциональности на портале размещение новых данных приостановлено), архив остается доступным и может быть полезен для исторических сравнений⁷. Этот источник не имеет аналогов, однако информация в нем остается агрегированной, поэтому ее анализ возможен лишь на уровне региона и типа преступления. Кроме того, стоит упомянуть коммерческие и справочно-правовые системы, такие как базы «Консультант+» и «Гарант», которые в том числе содержат информацию о судебных решениях, а также предоставляют льготный доступ для исследователей и преподавателей. Наконец, есть и ряд полностью коммерческих решений, специализирующихся на арбитражном процессе, например «Casebook» и «Caselook», имеющие API для доступа к данным и встроенные инструменты умного поиска. В отличие от вышеперечисленных систем, описываемый ниже подход основывается на работе с открытыми и непредобработанными данными.

С 2010 г. в России действует Федеральный закон от 22 декабря 2008 г. № 262-ФЗ «Об обеспечении доступа к информации о деятельности судов в Российской Федерации». Закон провозглашает «открытость и доступность информации о деятельности судов»⁸ и «свободу поиска, получения, передачи и распространения информации о деятельности судов»⁹. Перечень данных, подлежащих публикации, содержащийся в 14-й статье 262-ФЗ, достаточно широк и включает различные типы данных: от общих данных судебной статистики до отдельных текстов судебных решений.

Агрегированные данные судебной статистики в России публикуются на сайте Судебного департамента при Верховном суде РФ¹⁰ (далее — Судебный департамент), где представлены сводные статистические сведения о деятельности федеральных судов общей юрисдикции, а также о состоянии судимости, включая данные о числе привлеченных к уголовной ответственности и их социально-демографических признаках, видах наказания, обстоятельствах преступления. Аналогичная информация на уровне отдельных регионов представлена на сайтах региональных управлений при Судебном департаменте. Официальная статистика публикуется

⁷ Показатели преступности России // Генеральная прокуратура Российской Федерации. URL: <http://crimestat.ru/> (дата обращения: 16.12.2024).

⁸ Федеральный закон «Об обеспечении доступа к информации о деятельности судов в Российской Федерации» от 22.12.2008 № 262-ФЗ (последняя редакция) // КонсультантПлюс. URL: https://www.consultant.ru/document/cons_doc_LAW_82839/ (дата обращения: 18.02.2024).

⁹ Там же.

¹⁰ Данные судебной статистики // Судебный департамент при Верховном Суде Российской Федерации. URL: <http://cdep.ru/?id=79> (дата обращения: 18.02.2024).

в агрегированном виде (дела группируются по разным основаниям — например, по статье или группе статей УК РФ), что существенно ограничивает возможности для проведения исследований. Кроме того, анализ способов сбора статистических данных о преступности в России [Шклярчук, 2014; Шклярчук и др., 2015] указывает на существенные ограничения для их использования в исследовательских целях ввиду различных искажений, в целом свойственных данному источнику данных и рассмотренных нами выше.

Единственный доступный источник данных о судебной практике в РФ на дезагрегированном уровне — тексты судебных решений, публикация которых также стала обязательной для всех российских судов после вступления в действие 262-ФЗ. Тексты приговоров всех судов России различных уровней собираются в открытой ГАС «Правосудие»¹¹. Исключением является Москва — тексты приговоров московских судов публикуются отдельно на сайте Мосгорсуда¹². Кроме того, тексты арбитражных дел также публикуются в Картотеке арбитражных дел, имеющей свой программный интерфейс (API)¹³, что упрощает доступ исследователей к этим данным. Примером использования данного источника является работа В. В. Волкова, Д. А. Скугаревского, Р. К. Кучакова, в которой анализировались результаты арбитражных споров участников с различным судебным опытом [Volkov, Skougarevskiy, Kuchakov, 2023]. В нашем исследовании мы фокусируемся на анализе текстов приговоров по уголовным делам, хотя используемый нами подход может быть применен и для работы с гражданскими и административными делами.

Несмотря на то что статья 15 262-ФЗ устанавливает конкретные сроки публикации судебных решений (тексты судебных актов — не позднее месяца после дня их принятия, приговоров — не позднее месяца после дня их вступления в силу, судебных актов арбитражных судов — не позднее следующего дня после дня их принятия), на практике суды не всегда соблюдают эти сроки: для гражданских и уголовных дел данное требование не соблюдали 13—15 % судов, для административных дел эта доля достигала 45 % по состоянию на 2011 г. [Поздняков, 2011b]. Мониторинг, проведенный ИПП при ЕУ СПб годом позднее, показал, что скорость публикации в среднем увеличилась на несколько процентов, однако доля судов, которые в принципе игнорировали требования закона о публикации текстов судебных решений, увеличилась [Поздняков, 2012], при этом зачастую проблема доступа к этой информации заключается в том, что соответствующий раздел сайтов судов просто не работает [Поздняков, 2011a].

Правовые и технические ограничения работы с открытыми текстами приговоров

Прежде чем перейти к вопросу о сборе и предобработке судебных решений, необходимо рассмотреть нормативно-правовые ограничения, предусмотренные частями 3—5 статьи 15 № 262-ФЗ. Из текстов решений изымаются персональные

¹¹ Поиск по делам и судебным актам // ГАС РФ «Правосудие». URL: <https://bsr.sudrf.ru/bigs/portal.html> (дата обращения: 18.02.2024).

¹² Поиск по судебным делам // Официальный портал судов общей юрисдикции г. Москвы. URL: <https://mos-gorsud.ru/search> (дата обращения: 18.02.2024).

¹³ Картотека арбитражных дел // Pravo Tech. URL: <https://kad.arbitr.ru/> (дата обращения: 18.02.2024).

данные фигурантов, информация, составляющая медицинскую тайну или тайну личной переписки, а тексты по некоторым категориям дел (например, связанным с государственной тайной, преступлениями против половой неприкосновенности в отношении несовершеннолетних, акты с информацией об усыновлении или удочерении) запрещены к публикации.

Кроме того, существует практика частичной или избыточной деперсонализации приговоров, затрудняющая извлечение части данных, исключение которых не предписывается законом напрямую. По опыту работы с данными мы знаем, что из текстов приговоров может изыматься информация о гражданстве обвиняемого, его образовании и таких специфических особенностях совершения преступления, как, например, марка машины, ставшей причиной ДТП. В редких случаях в текстах может отсутствовать и срок вынесенного приговора.

Таким образом, исследователи должны учитывать ограничения, связанные с правовым регулированием и деперсонализацией текстов. Если часть информации изымается из документов в соответствии с нормами закона, то другая часть — в соответствии с внутренней практикой работы конкретных судов, которая может сильно различаться. Последнее означает, что при построении объяснительных моделей важно принимать во внимание фиксированные эффекты на уровне отдельного суда и региона.

Сбор и предобработка судебных приговоров

Тексты судебных приговоров и другая дополнительная информация по соответствующим уголовным делам могут быть собраны как с отдельных сайтов судов общей юрисдикции, так и с сайтов-агрегаторов судебных дел. Как было упомянуто, официальным агрегатором судебных решений в России выступает ГАС «Правосудие», которая должна быть дополнена официальным порталом судов общей юрисдикции города Москвы (далее — сайт Мосгорсуда). Однако с конца февраля 2024 г. на портале ГАС «Правосудие» перестал работать поиск по делам и текстам приговоров «в связи с модернизацией сайтов судов общей юрисдикции». По этой причине исследователи могут полагаться только на сайты отдельных судов, а также на неофициальные агрегаторы судебных решений (например, интернет-ресурсы «СудАкт»¹⁴ и «СудебныеРешения.рф»¹⁵). Однако, поскольку последние не являются официальными ресурсами, полнота собранных на их основе данных остается под вопросом. Кроме того, рубрикаторы поиска на указанных агрегаторах менее подробны, чем в ГАС «Правосудие» и на сайте Мосгорсуда, из-за чего формирование выборок дел по интересующим исследователя основаниям, например году вынесения приговора, не всегда оказывается возможным. В дальнейшем при описании процесса сбора приговоров мы фокусируемся на решении, предполагающем комбинирование данных из ГАС «Правосудие» (до момента приостановки его работы) и с сайта Мосгорсуда.

Тексты приговоров собираются с помощью техник веб-скрейпинга, поскольку ни один из доступных ресурсов не предоставляет возможности экспорта найден-

¹⁴ Судебные и нормативные акты РФ — Главная страница // Судебные и нормативные акты РФ. URL: <https://sudact.ru/> (дата обращения: 18.02.2024).

¹⁵ Судебные решения РФ // Павел Нетупский ООО «ПИК-пресс». URL: <https://судебныерешения.рф/> (дата обращения: 18.02.2024).

ных документов. В своей деятельности для сбора данных мы создали два алгоритма, для сайта Мосгорсуда и ГАС «Правосудие» соответственно, поиск и сбор дел в которых осуществляются по запросу, включающему статью УК РФ, год вынесения приговора и уровень суда. Программные скрипты для сбора данных опубликованы в открытом доступе¹⁶.

Отметим, что собственный скрейпер для сбора судебных дел по любой статье разработал также проект «Если быть точным»¹⁷. В нем, в отличие от наших алгоритмов, используется не ГАС «Правосудие», а отдельные сайты судов всех уровней, в связи с чем сам сбор данных занимает значительное количество времени и оказывается не чувствителен к фактическому объему подходящих под запрос исследователя данных. Одновременно с этим к преимуществам скрейпера «Если быть точным» можно отнести удобный и простой интерфейс, не требующий от пользователя навыков программирования. В условиях, пока ГАС «Правосудие» находится в процессе модернизации, вышеупомянутый скрейпер является единственным открытым и готовым решением по сбору информации с сайтов судов.

Собранным текстам приговоров требуется предварительная обработка для подготовки к извлечению необходимых характеристик из-за присутствия нестандартных знаков, множественных пробелов и ошибок квалификации. Кроме того, в данных могут попадаться дела с несколькими подсудимыми, а также обвинение может предъявляться сразу по нескольким статьям, что необходимо учитывать для формирования корректных выборок дел и их последующего анализа.

Предобработку текстов можно разбить на несколько этапов. На первом необходимо определить итоговую квалификацию преступления, так как квалификация в карточке дела не всегда соответствует финальной. Для этого требуется найти в тексте приговора его резолютивную часть и определить ту статью, по которой вынесено решение подсудимому, что можно сделать с помощью регулярных выражений (подробнее о них — в следующем разделе). Затем нужно определить количество подсудимых и наличие дополнительного состава преступления. Для этого все встречающиеся в наборе данных пункты и статьи обвинения можно отсортировать по максимальному наказанию, после чего в качестве основного состава выбрать статью с наибольшим потенциальным наказанием, а все остальные пункты обвинения классифицировать как дополнительный состав преступления. Количество подсудимых можно определить по числу фамилий и имен, встречающихся в списке обвиняемых.

На втором этапе предобработки тексты очищаются от сторонних символов (например, знаков html-разметки, использующейся в ГАС «Правосудие») и исправляются опечатки (множественные пробелы или, наоборот, отсутствие пробелов до и после использования знаков препинания), которые в дальнейшем могут затруднить корректное извлечение информации из текстов.

Третий этап предобработки опционален и необходим в случае, если в дальнейшем планируется применять методы машинного обучения (подробнее об этом — в следующем разделе). Этот этап предполагает приведение массива текстов к виду

¹⁶ См. подробнее: GAS-Mosgorsud-scrappers. URL: <https://github.com/vydevyatnikov/GAS-Mosgorsud-scrappers> (дата обращения: 18.04.2025).

¹⁷ SUDRF Scraper // Tochno-st. URL: <https://github.com/tochno-st/sudrfscraper> (дата обращения: 18.02.2024).

матрицы «документы-токены» и включает ряд стандартных шагов [Grimmer, Stewart, 2013]: удаление небуквенных символов из текстов, приведение текстов к нижнему регистру, лемматизацию слов (приведение их к словарной форме), а также удаление стоп-слов (часто встречающихся слов, не несущих смысловой нагрузки или слишком распространенных для конкретного эмпирического объекта, например, в случае анализа приговоров это слова «приговор», «российский», «федерация», с которых начинается любой приговор).

Извлечение информации из судебных приговоров

В настоящем разделе мы опишем основные техники извлечения информации из текстов судебных приговоров. В результате этого процесса исходный массив текстов трансформируется в пригодный для анализа массив количественных данных с наблюдениями и характеристиками этих наблюдений (переменными). В зависимости от исследовательского вопроса, целевых переменных и особенностей совершенного преступления (например, числа потерпевших) наблюдениями в итоговом массиве могут выступать разные сущности: само дело (преступление), отдельный обвиняемый, отдельный потерпевший и т. д. В дальнейшем для простоты изложения речь пойдет о ситуации, когда преступление совершено единолично и в нем один потерпевший — в таком случае наблюдением может служить само судебное дело. В качестве иллюстрации к излагаемым подходам мы воспользуемся недавним исследованием гендерных различий в приговорах за убийство по ст. 105 ч. 1 УК РФ в России [Zhuchkova, Kazun, 2023]. В этой работе с помощью описываемых далее подходов из текстов приговоров были извлечены следующие переменные: срок приговора, пол обвиняемого, пол потерпевшего, характер отношений между обвиняемым и потерпевшим (интимные партнеры, родственники, другое), наличие малолетних детей у обвиняемого, степень рецидива, степень признания вины, наличие иных смягчающих иотягчающих обстоятельств. Для анализа были автоматически обработаны тексты 20 531 приговора, а полученные из текстов данные опубликованы в открытом доступе¹⁸.

Необходимая информация из текстов судебных приговоров может извлекаться двумя путями: с помощью поиска релевантных фрагментов текста на основе специальных правил (rule-based approach) и с помощью построения моделей машинного обучения с учителем (machine learning approach). Базовый алгоритм действий по извлечению информации из текста состоит в том, чтобы сначала попробовать подход, основанный на правилах, поскольку он более прозрачный и менее трудозатратный. Если такой алгоритм не сработал, следует использовать машинное обучение. Наконец, техники могут комбинироваться для достижения большей точности извлекаемой информации.

Первая техника, основанная на правилах, применяется в случае, когда определенному значению искомой переменной соответствует конкретный фрагмент текста в приговоре, причем вариантов формулировок, по которым можно найти нужный фрагмент текста, ограниченное количество. С учетом высокой стандар-

¹⁸ См. подробнее: Zhuchkova S., Kazun A. Replication Data for: Exploring Gender Bias in Homicide Sentencing: An Empirical Study of Russian Court Decisions Using Text Mining // Harvard Dataverse. 2023. <https://doi.org/doi:10.7910/DVN/OYEPAA>.

тизированнойности структуры и языка приговоров это основной способ извлечения информации из них.

Правила, по которым ведется поиск релевантного фрагмента текста, могут быть простыми и сложными. Под простыми правилами мы понимаем случаи, когда проверяется наличие или отсутствие определенных формулировок в тексте без введения дополнительных условий. Например, пол обвиняемого может быть определен по наличию в приговоре слов «подсудимый» или «подсудимая» (в разных падежах). С помощью простых правил в рассматриваемом исследовании из текста приговоров была извлечена информация о поле обвиняемого, поле потерпевшего, наличии малолетних детей у обвиняемого.

Под сложными правилами мы понимаем случаи, когда поиск нужного фрагмента текста проходит в несколько этапов, в рамках которых последовательно проверяется несколько условий. Примеры переменных, которые требуют применения сложного правила, — наличие в деле смягчающих или отягчающих обстоятельств, эти переменные используются в качестве контрольных при изучении роли экстралегальных факторов судебных решений. Хотя формулировки, с помощью которых можно обнаружить упоминание таких обстоятельств, стандартны («обстоятельствами, смягчающими/отягчающими наказание...»), в тексте может идти речь о том, что суд их в итоге таковыми не признал. Одно из возможных решений, действующее сложное правило, может состоять в том, чтобы вначале находить упоминание смягчающих или отягчающих обстоятельств в тексте, а затем проверять, что предложение, в котором они упоминаются, не содержит выражений «судом не признается», «суд не принимает во внимание», «суд не учел» и проч. Срок приговора, степень рецидива, степень признания вины, наличие иных смягчающих и отягчающих обстоятельств — каждая из этих переменных в рассматриваемом исследовании была получена с применением сложных правил. Примеры правил представлены в [Приложении 2](#). Отметим, что всегда речь идет не о поиске статичного, неизменяемого текста, а об использовании регулярных выражений — специальных конструкций, которые наряду с привычным текстом могут содержать и метасимволы, позволяющие учесть возможные вариации в окончаниях, регистре слов, наличие каких-то дополнительных слов или символов между искомыми и т. д.

Вторая техника, основанная на машинном обучении с учителем, применяется для определения значения тех переменных, выделить ограниченный набор формулировок для которых оказывается ресурсозатратно или невозможно, а также в случае, когда для понимания смысла важны не отдельные фрагменты текста, а весь приговор целиком. В исследовании, взятом в качестве иллюстрации, таким образом определялся характер отношений между обвиняемым в убийстве и жертвой: были ли они интимными партнерами, родственниками или между ними были иные отношения. В тексте приговора нужная информация может быть описана разными формулировками: «он сожительствовал с ФИО13», «у него с потерпевшей были приятельские отношения», «с ранее ему не знакомым ФИО1» и т. п. Составить правило, которое бы учитывало все варианты таких фраз, невозможно, а сузить правило до поиска отдельных слов «сожительница», «мать» и проч. было бы ненадежно, поскольку указанные лица могут упоминаться в приговоре в разных ролях и контекстах. Обращение к моделям машинного обучения в дан-

ном случае предполагает, что случайная выборка приговоров должна быть прочитана вручную, значения целевой переменной для этой части приговоров должны быть закодированы человеком — и на этой части приговоров будет обучаться и тестироваться модель. Модель учится обнаруживать закономерности между словами в приговоре и меткой, которую проставил человек, причем эти закономерности могут быть довольно сложными, а итоговое решение строится исходя из содержания всего приговора целиком. Для оставшейся части приговоров — тех, что не были прочитаны человеком — значения переменной определяются с помощью предсказаний построенной модели. Перечень конкретных методов машинного обучения, которые могут применяться для описанной задачи классификации текстов, довольно широк и включает в себя регрессионные модели, деревья решений, ансамблевые методы, нейросети и проч. (см. например, [Gasparetto et al., 2022]). В исследовании, взятом в качестве примера, наиболее высокую точность предсказаний показал градиентный бустинг на деревьях решений. Классический алгоритм действий для выбора конкретного метода включает в себя тестирование разных методов (и параметров внутри каждого метода) и сравнение показателей точности прогнозов с последующим выбором наиболее точной и стабильной (не переобученной) модели.

Кроме методов классического машинного обучения, в последние годы получили широкое распространение и глубокие нейронные сети. Популярные алгоритмы, такие как BERT или GPT-модели, способны эффективно учитывать контекст и многозначность слов, обеспечивая высокое качество классификации и извлечения информации из текстов (см. обзор на эту тему [Казун, 2024]). Эти методы, будучи обученными на размеченных данных, требуют достаточного количества наблюдений (десятки тысяч текстов, если необходима высокая точность), а также тщательного контроля качества разметки и использования метрик оценки (таких как F1-score, точность, полнота). Впрочем, если у исследователей нет задачи добиться 99-процентной точности, может быть достаточно и тысячи размеченных вручную решений. Сопоставимые с этим объемы выборок для обучения моделей использовались и в описанных ранее примерах исследований: 1500 дел было закодировано вручную в исследовании гендерных различий в приговорах за убийство [Zhuchkova, Kazun, 2023], 1710 дел с убийствами женщин использовались для обучения модели определять в деле признаки домашнего насилия в проекте «Алгоритм света»¹⁹.

Две описанных техники — с привлечением правил и моделей машинного обучения — можно комбинировать для достижения высокой эффективности и точности извлечения признаков. Именно таким образом, например, решается задача извлечения информации об участниках судебного процесса, такой как их ФИО и роль в процессе (например, адвокат может выступать на стороне обвиняемого или жертвы). Эта классическая задача извлечения именованных сущностей NER (Named Entity Recognition), дополненная необходимостью вместе с ФИО определить роль участника судебного процесса, в силу высокой структурированности и клишированности текстов приговоров может быть эффективно решена с по-

¹⁹ Алгоритм света // Консорциум женских НПО. URL: <https://algorithmsveta.com/> (дата обращения: 18.02.2024).

мощью предобученной NER-модели Natasha²⁰ и правил. На наших данных точность такого решения составила примерно 94 %. Использование предобученной коммерческой GPT-модели для уточнения результатов позволило увеличить точность до 99 %. Такой подход дает возможность одновременно и сэкономить время на разметку данных, и без больших финансовых затрат получить высокую точность извлечения признаков.

Кроме извлечения непосредственно из текстов приговоров, информация по отдельным переменным может быть получена и с привлечением других источников данных. Так, при сборе информации о приговорах с ГАС «Правосудие» и сайта Мосгорсуда уместно включать в выгрузку не только сами тексты приговоров, но и данные из карточек дел — например, год приговора, регион, суд, ФИО судьи и др. Часть данных может быть использована в исходном виде, часть может послужить основой для определения других характеристик (например, на основе окончания фамилии или отчества судьи можно определить пол судьи). Перечисленные переменные также использовались в регрессии в рассматриваемом исследовании в качестве контрольных (год, регион, пол судьи) и кластерных (суд) переменных.

Выводы и заключение

Настоящую статью следует рассматривать как представление и общую инструкцию к использованию нового источника данных для эмпирических исследований права — текстов судебных приговоров. На основе имеющегося опыта мы описали, как выгрузить, предобработать и трансформировать тексты приговоров в привычный исследователям массив количественных данных, пригодных для анализа. Предложенный подход открывает широкие перспективы для социологических исследований, направленных на изучение взаимосвязи между результатами и качеством правоприменения и различными социальными факторами.

По сравнению с иными доступными данными, данные из судебных приговоров обладают рядом существенных преимуществ. Во-первых, она дезагрегированы. В отличие от данных официальной судебной статистики, которые публикуются в сгруппированном до статьи или группы статей УК РФ виде, данные из приговоров позволяют отвечать на исследовательские вопросы, касающиеся характеристик отдельных дел или участников дел (обвиняемых, потерпевших, судей и т. д.). Среди таких исследовательских вопросов — выявление предвзятости (bias) в судебных решениях, типичных обстоятельств совершения тех или иных преступлений, выявление различий между группами преступников (например, женщинами и мужчинами). Во-вторых, это содержательно насыщенные данные: за счет наличия в тексте приговора подробного описания обстоятельств преступления, мотивов и последствий его совершения, показаний свидетелей и следствия из него можно извлекать такие параметры, которые не фиксируются в ведомственной статистике [Kovalchuk et al., 2022] и редко исследуются даже за рубежом, например, характеризующие поведение потерпевшего или обстоятельства, случившиеся задолго до совершения преступления. Кроме того, данные из текстов приговоров можно дополнить данными из сторонних источников, описывающих суд,

²⁰ Natasha // Alexander Kukushkin. URL: <https://github.com/natasha/natasha> (дата обращения: 18.02.2024).

судью, регион, предыдущие или последующие этапы рассмотрения дела, упоминание дела в СМИ и проч., таким образом еще в большей мере расширив пул потенциальных исследовательских вопросов. В-третьих, это данные большого объема, потенциально приближенные к генеральной совокупности, что обеспечивает высокую статистическую мощность последующего анализа и позволяет изучать даже относительно редко встречающиеся феномены (например, детерминанты получения оправдательного приговора или особенности работы судов присяжных).

В России пока нет альтернативных источников данных о преступлениях, которые бы совмещали в себе все перечисленные преимущества и находились бы в открытом доступе. Максимально похожи по своим свойствам оцифрованные данные из судебных карточек, которые использовались в ряде исследований ИПП при ЕУ СПб [Волков, 2014; Четверикова, 2014; Volkov, 2016], однако эти данные были получены исследователями по запросу и тиражирование данной практики затруднительно. Использование похожих по свойствам данных, находящихся в открытом доступе, более эффективно для целей развития в России эмпирических исследований права.

Одновременно с этим при обращении к текстам судебных приговоров необходимо учитывать ограничения, вытекающие из характера этих документов и методов, которые могут быть использованы для их обработки. Во-первых, применение методов автоматизированного анализа текстов не гарантирует 100-процентную точность получаемых на выходе данных [Alcántara Francia, Nunez-del-Prado, Alatrísta-Salas, 2022]. При использовании подхода, основанного на правилах, можно не учесть отдельные формулировки и из-за этого некорректно определить значения переменных, а модели машинного обучения, если они не переобучены, то есть не подогнаны под обучающую выборку, в социальных исследованиях в принципе не способны дать абсолютно точного прогноза. Тем не менее наш опыт применения методов автоматизированного анализа текстов позволяет говорить о возможности добиться достаточно высокой точности извлечения данных из судебных приговоров — на уровне не менее 90 %, что существенно превышает средние метрики при использовании других источников текстовых данных в социальных науках.

Во-вторых, при всей насыщенности текстов приговоров некоторые, в том числе базовые, характеристики извлечь из них в полном объеме невозможно, что актуализирует проблему пропусков в данных [Wu et al., 2022]. Публикуемые российские судебные акты должны быть деперсонализированы, то есть персональные данные подсудимых должны быть либо изъяты, либо заменены. На практике это требование исполняется неоднозначно: некоторые суды удаляют информацию избыточно, из-за чего теряются важные детали, необходимые для анализа; некоторые же суды публикуют тексты почти без искажений²¹ [Поздняков, 2011b; Поздняков, 2012]. Это ограничение, например, не позволяет включить в анализ одну из важнейших характеристик обвиняемого — его возраст. Зарубежные исследователи предлагают решать проблему с отсутствием необходимых социально-демографических характеристик, опираясь на прокси-переменные, доступные

²¹ В том числе вовсе не выполняя требование о деперсонализации данных, что можно заметить в том числе по приведенным в приложении примерам.

в текстах, например, этничность подсудимого можно с определить по имени, если оно не изъято из приговора [Pina-Sánchez, Roberts, Sferopoulos, 2019], однако такой подход не может полностью решить проблему с пропусками в данных. В российском контексте кроме требования изъятия персональных данных из приговоров проблему пропусков усугубляет возможность рассмотрения дела в особом порядке, то есть без проведения судебного разбирательства. Приговоры по таким делам существенно менее информативны, чем дела, рассматриваемые в общем порядке: в случае обвинительных приговоров тексты не будут содержать доказательств, на которых основано решение суда [Четверикова, Титаев, 2019], и именно из этой части приговора можно извлечь информацию об экстралегальных факторах. Кроме того, как отмечено ранее, отдельные дела запрещены к публикации.

В-третьих, открытым остается вопрос о том, все ли приговоры, которые должны быть опубликованы согласно закону, действительно публикуются, — и если нет, то существуют ли какие-то систематические различия между опубликованными и неопубликованными приговорами, которые должны быть учтены при интерпретации получаемых результатов. Имеющиеся сведения о потенциальной смещенности опубликованных приговоров фрагментарны и не могут быть признаны актуальными с учетом периода проведения соответствующих исследований. Так, в исследовании 2012 г. установлено, что более активная публикация приговоров свойственна районным судам, нежели областным, а также судам из административных центров по сравнению с судами из более отдаленных локаций [Поздняков, 2012].

Наконец, не вся представленная в судебном приговоре информация может быть достоверной, поскольку показания многих участников судебного процесса могут быть искажены в пользу собственной выгоды или из-за естественных ограничений человеческой памяти. Кроме того, отсутствие упоминания каких-либо интересующих исследователя характеристик в приговоре вовсе не означает, что и в реальности они не имели места. На наш взгляд, актуальность перечисленных ограничений во многом зависит от поставленных исследовательских вопросов, при этом многие из вопросов могут быть решены с использованием более однозначных переменных, вероятность искажения которых невелика. Часть из этих ограничений может быть преодолена в том числе за счет отказа от использования методов автоматизированной обработки приговоров и перехода к их ручному контент-анализу — такой подход хотя и требует больших ресурсов и сокращает потенциальный объем выборки, вполне уместен для изучения тех характеристик, которые проблематично извлечь из текстов автоматически.

Список литературы (References)

1. Веркеев А. М., Волков В. В., Дмитриева А. В., Кнорре А. В., Кудрявцев В. Е., Кузнецова Д. А., Кучаков Р. К., Титаев К. Д., Ходжаева Е. А. Как изучать жертв преступлений? // Мониторинг общественного мнения: экономические и социальные перемены. 2019. № 2. С. 4—31. <https://doi.org/10.14515/monitoring.2019.2.01>.
Verkeev A. M., Volkov V. V., Dzmityriyeva A. V., Knorre A. V., Kudryavtsev V. E., Kuznetsova D. A., Kuchakov R. K., Titaev K. D., Khodzhaeva E. A. (2019) How to Study Vic-

- tims of a Crime? *Monitoring of Public Opinion: Economic and Social Changes*. No. 2. P. 4—31. (In Russ.) <https://doi.org/10.14515/monitoring.2019.2.01>.
2. Волков В. В. Влияние социального статуса подсудимого на решение суда // Журнал социологии и социальной антропологии. 2014. Т. 17. № 4. С. 62—85. Volkov V. V. (2014) The Influence of Socio-Economic Status of Defendants on Court Decisions. *Journal of Sociology and Social Anthropology*. Vol. 17. No. 4. P. 62—85. (In Russ.)
 3. Волков В. В. Эмпирическая социология права в условиях междисциплинарного синтеза // Социологические исследования. 2017. № 4. С. 34—42. Volkov V. V. (2017) The Empirical Sociology of Law in the Context of Interdisciplinary Synthesis. *Sociological Studies*. No. 4. P. 34—42. (In Russ.)
 4. Волков В. В., Скугаревский Д. А., Титаев К. Д. Проблемы и перспективы исследований на основе Big Data (на примере социологии права) // Социологические исследования. 2016. № 1. С. 48—58. Volkov V. V., Skougarevskiy D. A., Titaev K. D. (2016) Problems and Prospects for Studies Based on Big Data (the Case of Sociology of Law). *Sociological Studies*. No. 1. P. 48—58. (In Russ.)
 5. Казун А. П. Может ли искусственный интеллект прогнозировать решения суда? Систематический обзор международных исследований // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 100—122. <https://doi.org/10.14515/monitoring.2024.5.2580>. Kazun A. P. (2024) Can Artificial Intelligence Predict Court Decisions? A Systematic Review of International Research. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 100—122. <https://doi.org/10.14515/monitoring.2024.5.2580>. (In Russ.)
 6. Кнорре А. В., Титаев К. Д. Преступность и виктимизация в России. Результаты всероссийского виктимизационного опроса (Аналитический обзор). СПб.: ИПП ЕУ СПб, 2018. Knorre A., Titaev K. (2018) Crime and Victimization in Russia. The Results of the All-Russian Victimization Survey (Analytical Review). St. Petersburg: EUSP Press. (In Russ.)
 7. Поздняков М. Л. Доступ к судебным актам судов общей юрисдикции: мониторинг применения Федерального закона от 22 декабря 2008 г. № 262-ФЗ «Об обеспечении доступа к информации о деятельности судов в Российской Федерации». СПб.: ИПП ЕУ СПб, 2011а. Pozdnyakov M. (2011a) Access to Judicial Acts of Courts of General Jurisdiction: Monitoring the Application of Federal Law No. 262-FZ Dated December 22, 2008 “On Ensuring Access to Information on the Activities of Courts in the Russian Federation”. St. Petersburg: EUSP Press. (In Russ.)
 8. Поздняков М. Л. Мониторинг открытости правосудия. СПб.: ИПП ЕУ СПб, 2011б. Pozdnyakov M. (2011b) Monitoring the Openness of Justice. St. Petersburg: EUSP Press. (In Russ.)

9. Поздняков М. Л. Организационные и структурные ограничения при доступе к судебным актам судов общей юрисдикции. СПб.: ИПП ЕУ СПб, 2012.
Pozdnyakov M. (2012) *Organizational and Structural Restrictions on Access to Judicial Acts of Courts of General Jurisdiction*. St. Petersburg: EUSP Press. (In Russ.)
10. Серебренников Д., Титаев К. Д. Динамика преступности и виктимизации в России 2018—2021 гг. Результаты второго виктимизационного опроса: аналитический обзор. СПб.: ИПП ЕУ СПб, 2022.
Serebrennikov D., Titaev K. (2022) *The Dynamics of Crime and Victimization in Russia 2018—2021. The Results of the Second Victimization Survey: An Analytical Review*. St. Petersburg: EUSP Press. (In Russ.)
11. Титаев К. Д. Предварительное заключение в российской уголовной юстиции: социологический анализ вероятности предварительного заключения и его влияния на решение суда // Экономическая социология. 2014. Т. 15. № 3. С. 88—113.
Titaev K. D. (2014) *Pretrial Detention in Russian Criminal Justice: Sociological Analysis of the Probability of Pretrial Detention and its Influence on Court Decisions*. *Economic Sociology*. Vol. 15. No. 3. P. 88—113. (In Russ.)
12. Титаев К. Д., Скугаревский Д. А., Кнорре А. В., Кудрявцев В. Е., Шклярук М. С. Манифест новой количественной криминологии «Уголовная политика с опорой на данные» // Вестник Санкт-Петербургского университета. Право. 2018. Т. 9. № 1. С. 91—101. <https://doi.org/10.21638/11701/spbu14.2018.107>.
Titaev K. D., Skougarevskiy D. A., Knorre A. V., Kudryavtsev V. E., Shklyaruk M. S. (2018) *A New Quantitative Criminology Manifesto: Evidence Based Criminal Justice Policymaking*. *Vestnik of Saint Petersburg University. Law*. Vol. 9. No. 1. P. 91—101. (In Russ.) <https://doi.org/10.21638/11701/spbu14.2018.107>.
13. Фридман Л. М. Взросление: исследования права и общества вступают в эксклюзивный клуб // Право и правоприменение в зеркале социальных наук: хрестоматия современных текстов / под ред. Э. Л. Панеях. М.: Статут, 2014. С. 184—200.
Friedman L. M. (2014) *Coming of Age: Law and Society Enters an Exclusive Club*. In: Paneyakh E. (ed.) *Law And Application of Law in the Mirror of Social Sciences: Digest of Modern Texts*. Moscow: Statut. P. 184—200. (In Russ.)
14. Четверикова И. В. Роль семьи, профессиональной карьеры и пола подсудимых при вынесении приговоров российскими судьями // Журнал социологии и социальной антропологии. 2014. Т. 17. № 4. С. 101—123.
Chetverikova I. V. (2014) *The Impact of Family, Professional Status and Gender of Defendants on Sentencing Decisions by Russian Judges*. *Journal of Sociology and Social Anthropology*. Vol. 17. No. 4. P. 101—123. (In Russ.)
15. Четверикова И. В., Титаев К. Д. Социологический анализ оптимального правоприменения: типичные случаи избыточной криминализации экономической деятельности в России // Вопросы экономики. 2019. № 3. С. 142—158. <https://doi.org/10.32609/0042-8736-2019-3-142-158>.

- Chetverikova I. V., Titaev K. D. (2019) Sociological Perspective on the Optimal Law Enforcement: Typical Cases of Overcriminalization of Economic Activity in Russia. *Voprosy Ekonomiki*. No. 3. P. 142—158. (In Russ.) <https://doi.org/10.32609/0042-8736-2019-3-142-158>.
16. Шклярук М. С. Правовая статистика: системный конфликт между знанием о преступности и учетом ведомственной работы. СПб.: ИПП ЕУ СПб, 2014.
Shklyaruk M. S. (2014) Legal Statistics: A Systemic Conflict between Knowledge about Crime and Accounting for Departmental Work. St. Petersburg: EUSP Press. (In Russ.)
 17. Шклярук М., Скугаревский Д., Дмитриева А., Скифский И., Бегтин И. Криминальная статистика: механизмы формирования, причины искажения, пути реформирования: исследовательский отчет. СПб.; М.: Норма, Центр независимых социальных исследований и образования, 2015.
Shklyaruk M., Skougarevskiy D., Dzmitryieva A., Skifsky I., Begtin I. (2015) Criminal Statistics: Mechanisms of Formation, Causes of Distortion, Ways of Reform: Research Report. St. Petersburg; Moscow: Norma, Center for Independent Social Research and Education.
 18. Юрченко Р. Н., Сакалов Б. З. О применении новых технологий в системе правосудия // Вестник Института законодательства и правовой информации Республики Казахстан. 2009. Т. 2. № 14. С. 89—95.
Yurchenko R., Sakalov B. (2009) On the Application of New Technologies in the Justice System. *Bulletin of the Institute of Legislation and Legal Information of the Republic of Kazakhstan*. Vol. 2. No. 14. P. 89—95.
 19. Addington L. A. (2008) Assessing the Extent of Nonresponse Bias on NIBRS Estimates of Violent Crime. *Journal of Contemporary Criminal Justice*. Vol. 24. No. 1. P. 32—49. <https://doi.org/10.1177/1043986207312936>.
 20. Ahl B., Sprick D. (2018) Towards Judicial Transparency in China: The New Public Access Aatabase for Court Decisions. *China Information*. Vol. 32. No. 1. P. 3—22, <https://doi.org/10.1177/0920203X17744544>.
 21. Alcántara Francia O. A., Nunez-del-Prado M., Alatrística-Salas H. (2022) Survey of Text Mining Techniques Applied to Judicial Decisions Prediction. *Applied Sciences*. Vol. 12. No. 20. P. 10200. <https://doi.org/10.3390/app122010200>.
 22. Aletras N., Tsarapatsanis D., Preoŧiuc-Pietro D., Lamps V. (2016) Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Computer Science*. Vol. 2. P. e93. <https://doi.org/10.7717/peerj-cs.93>.
 23. Ansari S., He N. (2015) Convergence Revisited: A Multi-Definition, Multi-Method Analysis of the UCR and the NCVS Crime Series (1973—2008). *Justice Quarterly*. Vol. 32. No. 1. P. 1—31. <https://doi.org/10.1080/07418825.2012.718355>.
 24. Ashley K. D., Brüninghaus S. (2009) Automatically Classifying Case Texts and Predicting Outcomes. *Artificial Intelligence and Law*. Vol. 17. No. 2. P. 125—165. <https://doi.org/10.1007/s10506-009-9077-9>.

25. Cao Y., Ash E., Chen D. L. (2020) Automated Fact-Value Distinction in Court Opinions. *European Journal of Law and Economics*. Vol. 50. No. 3. P. 451—467. <https://doi.org/10.1007/s10657-020-09645-7>.
26. Gasparetto A., Marcuzzo M., Zangari A., Albarelli A. (2022) A Survey on Text Classification Algorithms: From Text to Predictions. *Information*, Vol. 13, No. 2, : 83. <https://doi.org/10.3390/info13020083>
27. Grimmer J., Stewart B. M. (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*. Vol. 21. No. 3. P. 267—297. <https://doi.org/10.1093/pan/mps028>.
28. Hartel P., van Wegberg R., van Staalduinen M. (2023) Investigating Sentence Severity with Judicial Open Data: A Case Study on Sentencing High-Tech Crime in the Dutch Criminal Justice System. *European Journal on Criminal Policy and Research*. Vol. 29. No. 4. P. 579—599. <https://doi.org/10.1007/s10610-021-09503-5>.
29. Hipp J. R., Bates C., Lichman M., Smyth P. (2019) Using Social Media to Measure Temporal Ambient Population: Does it Help Explain Local Crime Rates? *Justice Quarterly*. Vol. 36. No. 4. P. 718—748. <https://doi.org/10.1080/07418825.2018.1445276>.
30. Kaur H., Choudhury T., Singh T. P., Shamoan M. (2019) Crime Analysis using Text Mining. In: *2019 International Conference on Contemporary Computing and Informatics (IC3I)*. Singapore: IEEE. P. 283—288. <https://doi.org/10.1109/IC3I46837.2019.9055606>.
31. Kovalchuk O., Banakh S., Masonkova M., Berezka K., Mokhun S., Fedchyshyn O. (2022) Text Mining for the Analysis of Legal Texts. In: *2022 12th International Conference on Advanced Computer Information Technologies (ACIT)*. Ruzomberok: IEEE. P. 502—505. <https://doi.org/10.1109/ACIT54803.2022.9913169>.
32. Loftin C., McDowall D. (2010) The Use of Official Records to Measure Crime and Delinquency. *Journal of Quantitative Criminology*. Vol. 26. No. 4. P. 527—532. <https://doi.org/10.1007/s10940-010-9120-8>.
33. Lundman R. J. (2003) The Newsworthiness and Selection Bias in News About Murder: Comparative and Relative Effects of Novelty and Race and Gender Typifications on Newspaper Coverage of Homicide. *Sociological Forum*. Vol. 18. No. 3. P. 357—386. <https://doi.org/10.1023/A:1025713518156>.
34. Metsker O., Trofimov E., Petrov M., Butakov N. (2019) Russian Court Decisions Data Analysis Using Distributed Computing and Machine Learning to Improve Law-making and Law Enforcement. *Procedia Computer Science*. Vol. 156. P. 264—273. <https://doi.org/10.1016/j.procs.2019.08.202>.
35. Moens M.-F. (2007) Summarizing Court Decisions. *Information Processing & Management*. Vol. 43. No. 6. P. 1748—1764. <https://doi.org/10.1016/j.ipm.2007.01.005>.

36. Nolan III J. J., McDevitt J., Cronin S., Farrell A. (2004) Learning to See Hate Crimes: A Framework for Understanding and Clarifying Ambiguities in Bias Crime Classification. *Criminal Justice Studies*. Vol. 17. No. 1. P. 91—105. <https://doi.org/10.1080/08884310420001679361>.
37. Ozkan T. (2019) Criminology in the Age of Data Explosion: New Directions. *The Social Science Journal*. Vol. 56. No. 2. P. 208—219. <https://doi.org/10.1016/j.soscij.2018.10.010>.
38. Paneyakh E. (2014) Faking Performance Together: Systems of Performance Evaluation in Russian Enforcement Agencies and Production of Bias and Privilege. *Post-Soviet Affairs*. Vol. 30. No. 2—3. P. 115—136. <https://doi.org/10.1080/1060586X.2013.858525>.
39. Pickett J. T., Mancini C., Mears D. P., Gertz M. (2015) Public (Mis)Understanding of Crime Policy: The Effects of Criminal Justice Experience and Media Reliance. *Criminal Justice Policy Review*. Vol. 26. No. 5. P. 500—522. <https://doi.org/10.1177/0887403414526228>.
40. Pina-Sánchez J., Grech D., Brunton-Smith I., Sferopoulos D. (2019) Exploring the Origin of Sentencing Disparities in the Crown Court: Using Text Mining Techniques to Differentiate between Court and Judge Disparities. *Social Science Research*. Vol. 84. Art. 102343. <https://doi.org/10.1016/j.ssresearch.2019.102343>.
41. Pina-Sánchez J., Roberts J. V., Sferopoulos D. (2019) Does the Crown Court Discriminate Against Muslim-named Offenders? A Novel Investigation Based on Text Mining Techniques. *The British Journal of Criminology*. Vol. 59. No. 3. P. 718—736. <https://doi.org/10.1093/bjc/azy062>.
42. Poulin D., Mowbray A. (2007) Free Access to Law and Open Source Software. In: St. Amant K., Still B. (eds.). *Handbook of Research on Open Source Software*. Hershey, PA: IGI Global. P. 373—381. <https://doi.org/10.4018/978-1-59140-999-1.ch029>.
43. Prasad N., Boughanem M., Dkaki T. (2024) Exploring Large Language Models and Hierarchical Frameworks for Classification of Large Unstructured Legal Documents. Springer. https://doi.org/10.1007/978-3-031-56060-6_15.
44. Rand M. (2006) The National Crime Victimization Survey: 34 Years of Measuring Crime in the United States. *Statistical Journal of the United Nations Economic Commission for Europe*. Vol. 23. No. 4. P. 289—301. <https://doi.org/10.3233/SJU-2006-23406>.
45. Schwartz J., Vega A. (2017) Sources of Crime Data. In: Teasdale B., Bradley M. S. (eds.) *Preventing Crime and Violence*. Cham: Springer International Publishing. P. 155—167. https://doi.org/10.1007/978-3-319-44124-5_14.
46. Shu D., Sen S. (2023) Analyzing Hong Kong's Legal Judgments from a Computational Linguistics point-of-view. arXiv. <https://doi.org/10.48550/arXiv.2305.02558> (accessed: 28.04.2025).

47. Turkanova V. (2023) Prospects for the Use of Artificial Intelligence and Machine Learning Algorithms for Effective Resolution of Civil Disputes. *Access to Justice in Eastern Europe*. Vol. 6. No. 2. P. 232—241. <https://doi.org/10.33327/AJEE-18-6.2-n000224>.
48. Volkov V. (2016) Legal and Extralegal Origins of Sentencing Disparities: Evidence from Russia's Criminal Courts. *Journal of Empirical Legal Studies*. Vol. 13. No. 4. P. 637—665. <https://doi.org/10.1111/jels.12128>.
49. Volkov V., Skougarevskiy D., Kuchakov R. (2023) Whither Repeat Players? Litigation Experience and Success in Court: Evidence from Russian Commercial Courts. *Social Science Research*. Vol. 112. P. 102875. <https://doi.org/10.1016/j.ssresearch.2023.102875>.
50. Wu X., Roberts M. E., Stern R. E., Liebman B. L., Gupta A., Sanford L. (2022) Augmenting Serialized Bureaucratic Data: The Case of Chinese Courts. *21st Century China Center Research Paper No. 2022—11*. URL: <https://doi.org/10.2139/ssrn.4124433>.
51. Zhuchkova S., Kazun A. (2023) Exploring Gender Bias in Homicide Sentencing: An Empirical Study of Russian Court Decisions Using Text Mining. *Homicide Studies*. <https://doi.org/10.1177/10887679231217159>.