

# Большие данные в юридическом образовании

Материалы открытого семинара  
из цикла «Как учить юристов»  
факультета права НИУ ВШЭ  
18 апреля 2023



ЦТЮО

Центр трансформации юридического образования  
факультета права НИУ ВШЭ

## Большие данные в юридическом образовании

Расшифровка материалов открытого семинара из цикла  
«Как учить юристов» (факультета права НИУ ВШЭ, 18 апреля 2023)

### Участники:



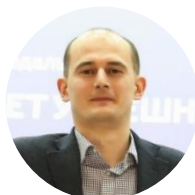
Дмитрий Текутьев, заместитель заведующего  
Центром трансформации юридического образования  
факультета права ВШЭ



Екатерина Безызвестных  
специалист по УМР ИТМО



Денис Савельев  
научный сотрудник ИПП ЕУСПб



Александр Чаплинский,  
старший научный сотрудник ИГМУ ВШЭ

© Участники, 2023.

Файл распространяется по лицензии Creative Commons «Attribution-NonCommercial-NoDerivatives» («Атрибуция — Некоммерческое использование — Без производных произведений») 4.0 Всемирная. Вы можете распространять это произведение при условии, что будете делать это бесплатно и не будете вносить изменения в текст. Если вы хотите использовать расшифровку иным способом, свяжитесь с нами.

# Содержание

Дмитрий Текутьев. Большие данные в юридическом образовании.....	6
Екатерина Безызвестных. Проектирование образовательных программ на основе данных .....	16
Денис Савельев. Цифровые следы в исследовании жизни права.....	20
Александр Чаплинский. Анализ сложности НПА: на стыке права и компьютерной лингвистики .....	25
Ответы на вопросы.....	31

## О нас

Это материал открытого семинара, который провел Центр трансформации юридического образования факультета права НИУ ВШЭ. Мы занимаемся производством новых курсов и образовательных программ, собираем и распространяем лучшие практики и развиваем преподавание права в России.

Наши приоритетные направления:

- Преподавание технологий юристам, программы legal + tech;
- Развитие письменной речи юристов;
- Развитие правовой грамотности у не-юристов.

Мы сотрудничаем с преподавателями, программами и целыми вузами. Собираем материалы, делимся своими наработками, организовываем встречи, проводим воркшопы и курсы для преподавателей. Следить за нашей работой проще всего в [телеграм-канале](#) или [на сайте](#).

Хотите работать с нами? Напишите на [transform@hse.ru](mailto:transform@hse.ru).

# Введение

Мы продолжаем серию открытых семинаров, которые мы назвали «Как учить юристов». Здесь мы обсуждаем проблемы российского юридического образования и современные тренды обучения юристов. Это уже пятая подобная встреча.

Две недели назад мы обсуждали использование искусственного интеллекта в преподавании права, и я очень рад, что это обсуждение вызвало большой интерес среди множества участников и просмотров на YouTube. Сегодня у нас тоже «высокотехнологичная» тема: мы поговорим об использовании Big Data (больших данных) в юридическом образовании. Для начала, по традиции, представлю участников нашей сегодняшней дискуссии.

Меня зовут Дмитрий Текутьев, я заместитель руководителя центра трансформации юридического образования и по совместительству доцент факультета права Вышки. Наши спикеры сегодня:

**Екатерина Анатольевна Безызвестных** – кандидат педагогических наук, специалист по учебно-методической работе, Университет ИТМО

**Денис Александрович Савельев** – кандидат юридических наук, научный сотрудник Институт проблем правоприменения при Европейском университете в Санкт-Петербурге

**Александр Владимирович Чаплинский** – кандидат юридических наук, старший научный сотрудник Института государственного и муниципального управления ВШЭ

Запись мероприятия доступна на [YouTube канале факультета права](#) и в телеграм-канале [«Юридическое образование»](#).

# Большие данные в юридическом образовании

Дмитрий Текутьев

Моё выступление состоит из двух частей. В первой части я поделюсь своим пониманием роли Big Data в подготовке юристов, а также своим опытом изучения такой дисциплины, как Empirical Legal Studies, и использования полученных знаний в преподавании права и проведении исследований. Во второй части я расскажу об исследовании трендов рынка юридического образования, которое мы с коллегами провели в прошлом году. Это, на наш взгляд, может быть одним из направлений использования Big Data, - я имею в виду концепцию «data driven learning», когда мы проводим анкетирование и опираясь на данные смотрим какие проблемы и потребности присутствуют в юридическом образовании в целом, в каком-то сегменте образования, или у конкретной группы или аудитории.

Что касается использования Big Data в правовых исследованиях, я думаю, что коллеги меня в этом поддержат: в России достаточно мало правовых исследований, где присутствует эмпирическая часть. Если мы посмотрим на ведущие западные журналы, в статьях по праву практически всегда присутствует эмпирическая часть, то есть часть, где некая правовая гипотеза проверяется на основе определенных данных. Плюс, есть достаточно много специализированных юридических журналов, которые посвящены эмпирическим исследованиям.

[К примеру, в Америке уже более 15 лет издается отдельный журнал, посвященный эмпирическим исследованиям в праве - Journal of Empirical Legal Studies.](#)

К сожалению, в России подобные исследования встречаются крайне редко. Зачастую, у нас правовые гипотезы или предложения по реформам действующего права основываются скорее на принадлежности автора к той или иной доктринальной позиции или школе (например, американист, германист, цивилист, хозяйственник) или на основании абстрактного представления исследователя о справедливости или эффективности предлагаемого регулирования. На мой взгляд, это большая проблема,

потому что такие исследования права далеко не всегда получаются объективными.

Большие данные помогают сделать исследование максимально объективным, в частности, исключить у автора когнитивные искажения, которым подвержены все люди, включая юристов. Знаменитая классификация Канемана разделяет мышление на систему 1 (основанную на интуиции, эмоциях и стереотипах) и систему 2 (более рациональную и аналитическую). Проверка гипотезы с использованием Big Data помогает максимально исключить нерациональные суждения в области права и активировать систему 2.

Это относится не только к научным исследованиям, но и к законодательной работе. Когда мы хотим ответить на какой-то правовой вопрос или решить правовую проблему, нам часто бывает полезно провести исследование данных и научно обосновать наши предположения. В противном случае, это может быть не научное утверждение, а просто предположение или стереотип. Законы – достаточно серьезная вещь, чтобы основывать их на предположениях, которые не подтверждены статистически и научно. Примеров применения такого подхода можно привести множество.

Я могу привести ряд примеров из своей области, корпоративного права, потому что мне это наиболее близко.

Скажем, если нас интересует эффект отведения в компании определенной структуры совета директоров, например, влияние наличия независимых директоров или наличия комитета по аудиту, прежде чем выступать с такой инициативой, нам важно понять, есть ли статистика, которая показывает, что определенная структура совета директоров влияет на прибыль компании, количество рисков и другие оцениваемые параметры.

Мы можем использовать статистику в динамике, чтобы ответить на этот вопрос.

Например, если у компании был стандартный совет директоров, а затем они переходят к независимым директорам, то мы можем сравнить периоды до и после появления независимых директоров и оценить динамику прибыли и других показателей.

Таким же образом мы можем оценивать эффекты регулирования в праве. Например, в стране принимается новое регулирование деятельности акционерных обществ, мы хотим посмотреть, как оно влияет на стоимость акций компаний. Мы анализируем динамику до и после введения регулирования и делаем определенный вывод. Кроме того, в судах юристы часто сталкиваются с необходимостью доказывать причинно-следственную связь. Знание основ статистик и социологии может помочь правильно

установить эту связь и избежать типичных ошибок – ошибок первого и второго рода, как не перепутать корреляцию и причинно-следственную связь и так далее.

Однако проблема состоит в том, что студентов юридических факультетов практически нигде не учат статистике, социологии и работе с данными, хотя это знание было бы полезно для них.

На Западе преподавание статистики для юристов является абсолютно нормальной практикой. Когда я обучался в магистратуре в Нидерландах, у нас был предмет «Эмпирические правовые исследования», который состоял из теоретической и практической частей. В практической части курса мы решали задачи по использованию правовой статистики в программе Stata и других программах. На теоретических занятиях мы изучали базовые понятия статистики и как их можно использовать в праве чтобы избежать неверных выводов и ошибок. Такие вещи как стандартное отклонение, нормальное распределение, закон больших чисел, динамический ряд, ошибка упущенной переменной и т.д. Для экономистов и социологов знание этого при проведении исследований является очевидным, но для юристов – далеко не всегда. По моему опыту, студенты юридических факультетов далеко не всегда даже знают, чем отличается среднее значение, мода и медиана. Поэтому я стараюсь посвятить хотя бы одно или две занятия статистике в рамках курса экономического анализа права, который я веду в Вышке. Также я стараюсь пропагандировать это направление среди студентов, чтобы они включали эмпирические части в свои курсовые и дипломные работы.

## Часто ли студенты делают подобные исследования?

Работы с эмпирическими компонентами встречаются довольно редко, в моей практике только одна работа за последние два года была выполнена с таким подходом. В основном, студенты неохотно занимаются такими исследованиями из-за отсутствия знаний и инструментария для этого.

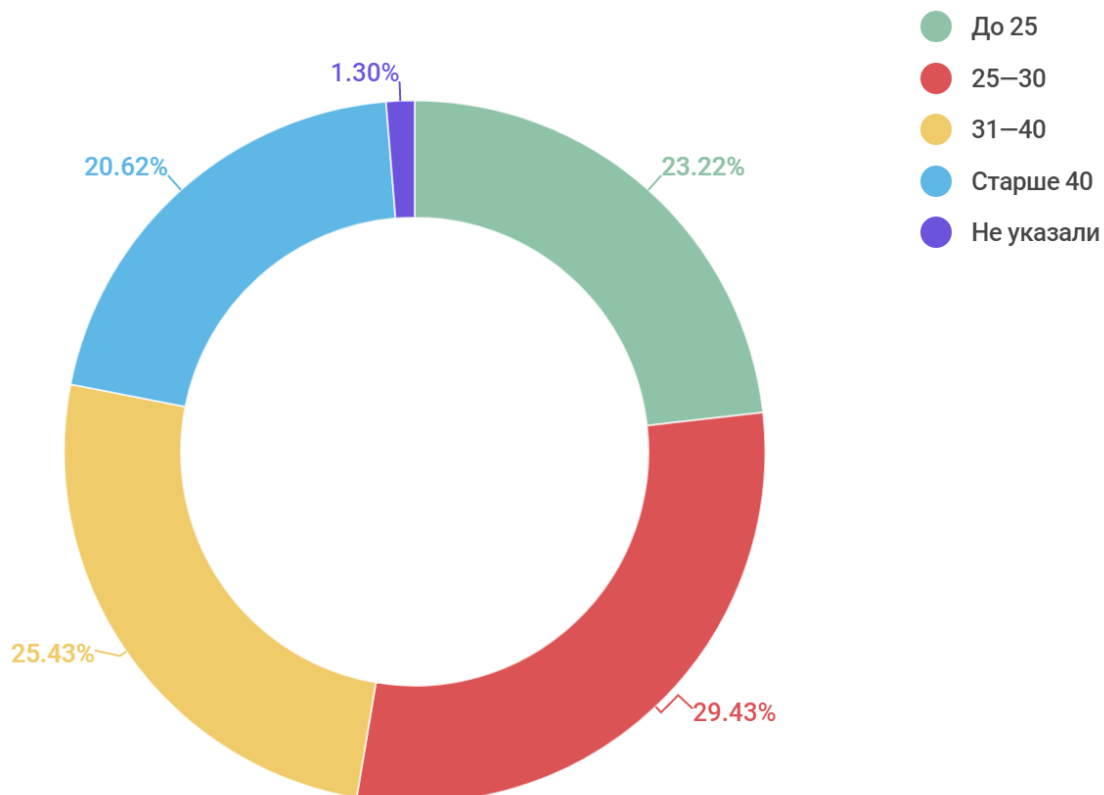
Однако на практике возникает следующая проблема: сложно оценить не юридическую часть таких работ, так как большинство тех, кто будет их проверять (рецензенты, члены экзаменационных комиссий), являются юристами и не всегда обладают компетенциями для грамотной оценки качества эмпирического исследования, выполненного студентом.

Возможным решением этой проблемы может стать приглашение социологов или специалистов по статистике на защиту дипломных работ, чтобы они могли оценить именно эмпирическую часть работы. Это может

помочь студентам получить более полноценную обратную связь и оценку своих исследований, а также привлечь больше внимания к значимости статистики и эмпирических исследований в области права.

## Исследование трендов рынка юридического образования

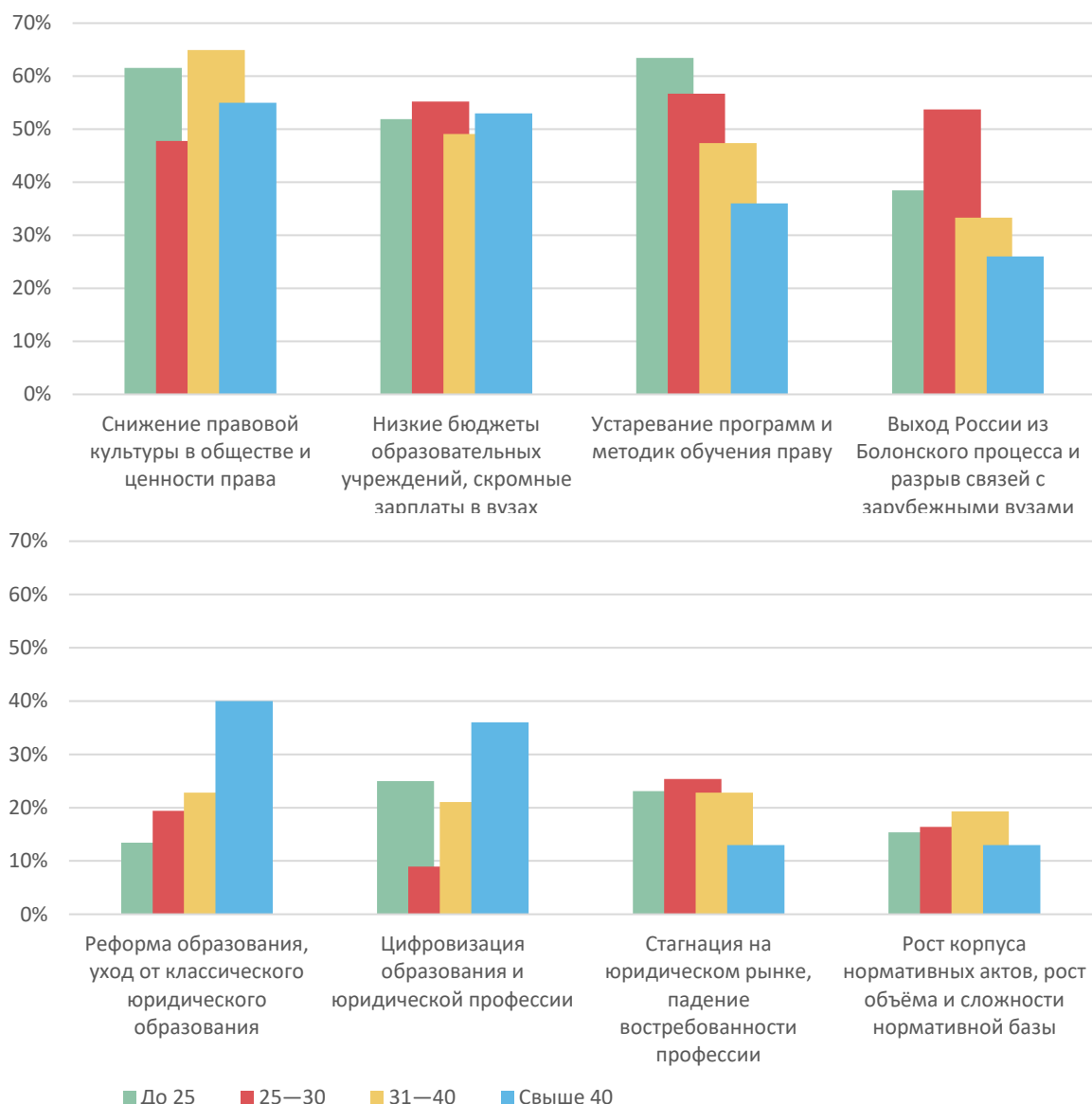
Теперь я перейду ко второй части своего выступления, к нашему исследованию трендов рынка юридического образования. Вкратце расскажу его предысторию. В декабре прошлого года Роман Янковский провел опрос в Facebook на тему «Будущее юридического образования и науки». В опросе приняли участие 230 человек, и, хотя мы понимаем, что подобная выборка может быть нерепрезентативной, она включала достаточно разнообразную аудиторию по возрасту и профессиональным характеристикам. Большинство участников опроса были практикующими юристами и преподавателями права, с некоторым перевесом в сторону молодых людей (до 31 года).



Опрос содержал 9 вопросов с вариантами ответов по системе multiple choice, и его результаты могут быть использованы для обсуждения трендов в развитии юридического образования и науки, служить отправной точкой для дискуссии и выработки стратегии развития данного направления среди студентов и преподавателей.

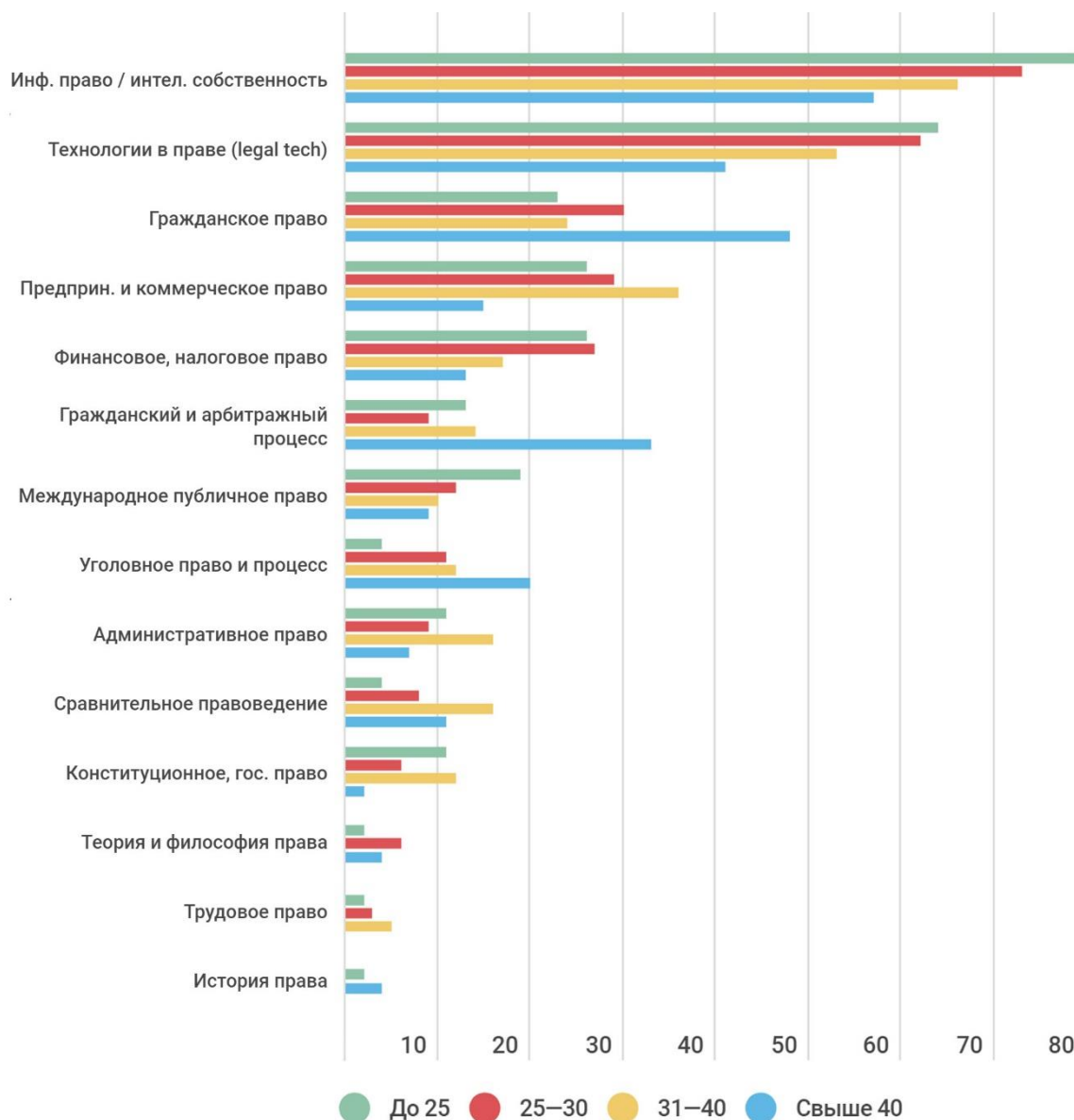
В частности, в результате опроса были выявлены ключевые вызовы для юридического образования, воспринимаемые респондентами с разными возрастными и профессиональными характеристиками.

К примеру, опрос показал, что молодые респонденты больше обеспокоены устареванием программ обучения, в то время как респонденты старшего возраста (31-40 лет) более негативно относятся к снижению правовой культуры в обществе. Практикующие юристы в большей степени озабочены снижением правовой культуры, а преподаватели – низкими бюджетами вузов.



В вопросе о том, какие специализации станут более актуальными через 10 лет, респонденты выделили IP, legaltech, а также традиционные частноправовые отрасли. При этом молодые респонденты до 25 лет гораздо чаще выбирали legaltech, в то время как респонденты среднего и старшего возраста больше отдавали предпочтение частноправовым отраслям. Респонденты старше 50 лет также достаточно часто называли уголовное право и уголовный процесс.

Среди наименее актуальных специализаций через 10 лет - история, теория и философия права, сравнительное правоведение, а также публично-правовые отрасли – конституционное, государственное и международное публичное право. Причем, именно молодые респонденты (до 25) чаще других называют историю, теорию и философия права, а также публичные отрасли права в качестве наименее актуальных.

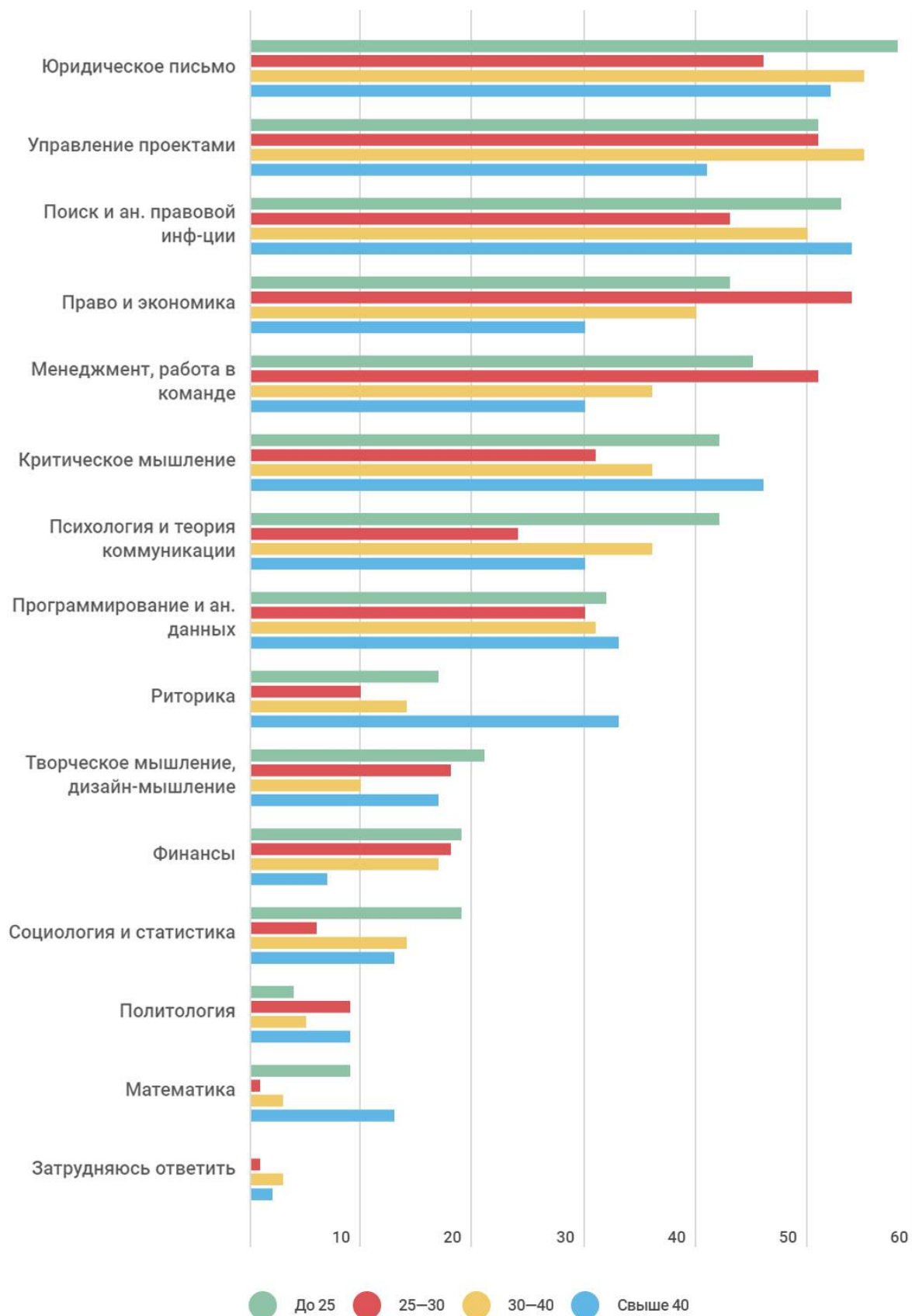


Среди наиболее вероятных характеристик юридического образования будущего респонденты отмечают появление новых форматов проведения занятий (в том числе онлайн), а также замену бумажных учебников онлайн пособиями (особенно этот ответ преобладает среди студентов). При этом в то, что юридическое образование полностью перейдет через 10 лет в онлайн верит крайне незначительный процент опрошенных. В целом, большинство респондентов считают, что после окончания ВУЗа будет работать около 50% студентов.

Что касается предметов, которые станут нормой через 10 лет, то здесь большинство выделяют юридическое письмо, поиск и анализ правовой информации, управление проектами, а также экономический анализ права и критическое мышление.



Мы считаем, что полученные результаты в целом отражают ожидания и предположения студентов, преподавателей и практикующих юристов относительно будущего юридического образования. Эти данные могут быть использованы для обсуждения и формирования стратегий развития юридического образования с учетом требований рынка труда и предпочтений студентов.



Мы опубликуем подробные результаты исследования в нашем Telegram-канале «Юридическое образование» и будем рады выслушать критику и предложения от профессионалов, в том числе социологов, чтобы улучшить нашу методологию и выводы.

## Какие выводы мы сделали?

Основываясь на результатах опроса, в первую очередь для своей практической деятельности, можно сделать следующие выводы:

1. Учитывая, что большинство респондентов считают юридическое письмо, поиск и анализ правовой информации, управление проектами, экономический анализ права и критическое мышление важными предметами в будущем, учебные заведения должны активно внедрять и развивать эти направления в своих образовательных программах.
2. Стоит уделить внимание внедрению онлайн-образования и замене бумажных учебников онлайн-пособиями, поскольку большинство респондентов ожидают таких тенденций в будущем.
3. Большинство респондентов ожидают роста междисциплинарных проектов и исследований, юридические вузы должны более активно сотрудничать с другими научными подразделениями и создавать совместные исследовательские программы.
4. Необходимо уделить внимание развитию компьютерных наук, социологии и экономики в юридических исследованиях, поскольку большинство респондентов считают, что методы этих наук будут использоваться чаще всего.

# Проектирование образовательных программ на основе данных

## Екатерина Безызвестных

Тема доказательного образования становится все более актуальной. Последние годы я активно внедряю Data Driven подход в практику проектирования. Также я занимаюсь обучением коллег, проектировщиков. Работая с данными и используя Data Driven подход в проектировании образовательных программ, мы можем сделать обучение более результативным, адаптированным к потребностям студентов и рынка труда, а также сделать образовательный процесс более прозрачным и обоснованным.

При проектировании курсов мы чаще всего сталкиваемся с тем, что нет ценности к аналитическим, исследовательским компетенциям. Наши коллеги не понимают зачем это нужно. Поэтому мы стараемся эту ценность к доказательному образованию, к анализу, сформировать.

Мы работаем с моделью ЭДИ, когда начинаем проектировать курс на основе данных. Она хороша тем, что в ней последовательно расписаны все этапы, нельзя что-то упустить:

1. Все начинается с анализа целевой аудитории, целей обучения, ресурсов образовательной среды. В таком подходе под ресурсами мы понимаем не только техническое оснащение, но и человеческие ресурсы.
2. Далее проектируем программы курсов. Мы уже понимаем какие есть возможности, вводные для реализации программы.
3. Следующий этап - разработка учебно-методических материалов. Например, методички, видео для онлайн курсов. Если мы используем LMS, то это все, что связано с загрузкой заданий в LMS. Мы опираемся на данные, собранные на предыдущих этапах, и используем их для создания качественных и актуальных материалов, адаптированных под потребности нашей аудитории и рынка труда.

4. Внедрение образовательного решения, запуск процесса.
5. Оценка.

На этапе анализа мы сталкиваемся с разными сильными сторонами, но я всегда рекомендую обращаться к слабым. Когда мы собираем данные, прежде всего стоит обращать внимание на эти места.

## С какими сложностями мы сталкиваемся?

Когда мы анализируем рынок конкурента, то, как правило, фокусируемся на лидерах. При этом мы обращаем меньше внимания на косвенных конкурентов. Поэтому при аналитике мы всегда рекомендуем коллегам обращать внимание не только на тех, кто на слуху, но и обязательно проанализировать регионы.

При анализе целевой аудитории мы рекомендуем использовать проверенные инструменты, например, «Карта эмпатии», метод персон, ценностное предложение. Однако, здесь есть такой нюанс, как слабая связь с работодателями. На этом этапе важно понимать какие новые профессии будут востребованы.

Образовательная среда, в которой реализуются образовательные программы, также оказывает значительное влияние на их успех и качество. Создание благоприятных условий для сотрудничества с бизнесом, индустрией, научными и технологическими партнерами, а также налаживание взаимодействия между участниками образовательного процесса, способствует стимулированию инноваций, разработке и внедрению новых образовательных решений.

Одним из ограничений в гипотезе является то, что очень много разных факторов влияет на метрики и показатели для оценки качества образовательных продуктов. Мы сталкивались с тем, что, когда проводим исследования среди нашей целевой аудитории, возникает проблема, связанная с небольшими выборками.

Двигаясь дальше, я хочу поделиться примером из одного из образовательных мероприятий, которые мы проводили - летней школы для преподавателей и методистов по проектированию обучения на основе данных. Участники формулировали гипотезы, приводили примеры изменений в продукте, описывали целевую аудиторию, метрики и результаты, а также планировали действия.

На этапе проектирования мы обращаем внимание на то, какие данные используем. Здесь мы анализируем все, что связано с целями обучения, мотивацией, вовлеченностью, опытом и форматом. Однако, в качестве слабых мест, часто сталкиваемся с неточно согласованными компонентами и методиками обучения. Сегодня существует множество авторских методик, и возникают вопросы о том, что является водителем и видным результативными показателями.

По поводу качества инструментов оценивания и сбора данных, мы используем различные качественные и количественные опросники, тестирование, а также более сложные инструменты. На этапе проектирования важно учесть способы расчета метрик и показателей, так как некоторые из них достаточно простые и понятные, но часто мы сталкиваемся с более сложными формулами. Бывает, что метрики необходимо подгонять и проектировать индивидуально для каждого продукта, учитывая его специфику.

Я хочу поделиться с вами одной из авторских методик педагогического дизайна, которую мы разработали. Здесь мы приводим примеры того, как можно применять те или иные методики на разных этапах проектирования, реализации обучения в контрольных точках и после обучения. На примере методик обратной связи и оценивания, мы устанавливаем разные этапы и методики, которые можно применять.

Следующий этап – разработка, который тоже очень важен. Здесь мы анализируем все, что связано с сервисами и платформами. В нашем университете, например, существует демократичная среда, где можно выбирать любые платформы для проведения занятий и консультаций, без единого обязательного решения. Каждый преподаватель может выбрать то, что ему нравится, и обсудить со студентами тот или иной сервис для дальнейшей работы. Это также связано с платформами открытого образования и различными LMS. Как я уже отметила, в университете очень демократичная среда и много возможностей для вариативности.

Переходим к этапу реализации. Здесь мы обсуждаем все, что связано с изменением образовательной среды, обратной связью и оцениванием. Приводим разные методики, которые мы используем для сбора данных. Если говорить о оценивании, примеры разных форматов включают самостоятельное оценивание, взаимное, экспертное и автоматическое. Вот тоже приведены примеры слабых мест, связанных с оцениванием, таких как проблемы в обработке и ошибки при применении инструментов. Очень важно обращать на них внимание.

Обратная связь также очень важна, и мы анализируем ее по всем нашим программам, собирая обратную связь на разных этапах, во время обучения и после обучения, а также отложенную обратную связь. Одной из основных проблем является непрозрачность применения результатов

обучения, то есть не всегда очевидно и понятно, как они используются в настоящее время или в будущем. Это также относится к обучению наших сотрудников и преподавателей, где иногда сложно отследить результативность в данный момент и, соответственно, принять решения о корректировке содержания программы или других аспектов.

Пример метрик, на которые мы обращаем внимание: удовлетворенность аспектами обучения, процент посещаемости курса, процент доходности курса, актуальность контента и такие важные составляющие, как безопасность и экологичность образовательного пространства.

Заключительным этапом является оценка, и здесь мы, конечно, анализируем все полученные данные, чтобы улучшить образовательные программы и подходы.

# Цифровые следы в исследовании жизни права

Денис Савельев

Право могут исследовать не только юристы, но и социологи, экономисты, политологи. Это действительно очень интересно и позволяет получить новые углы зрения на традиционные правовые проблемы и вопросы. Сейчас юристы только начинают открывать для себя доказательные методики и решения, основанные на данных.

Наше исследование – это такой междисциплинарный подход к исследованию права. В своем исследовании жизни права с помощью цифровых следов, мы стараемся применять различные методы анализа данных, которые позволяют нам получить дополнительную информацию о поведении людей и их взаимодействии с правовой системой. Это включает в себя изучение онлайн-поведения, анализ текстов законов и решений судов, а также применение статистических методов для анализа данных об исках и судебных разбирательствах. Таким образом, мы можем выявить закономерности, тенденции и проблемы, которые могут быть использованы для улучшения правовой системы и создания более справедливого и эффективного правопорядка.

При этом, мы стараемся сотрудничать с представителями разных профессий, так как именно междисциплинарный подход позволяет нам получить максимально полное и объективное представление о жизни права и его роли в обществе. Применение цифровых следов в исследовании жизни права открывает новые возможности для исследователей и практиков, позволяя им лучше понимать сложные взаимосвязи между правовой системой, обществом и индивидами. Это, в свою очередь, может способствовать созданию более эффективных и справедливых законов и механизмов правоприменения, а также формированию более грамотного и осведомленного общественного мнения о праве и его роли в жизни людей.

Хочу отметить, что исследование жизни права с помощью цифровых следов является важным и перспективным направлением научных исследований, и мы рады видеть, что все больше ученых и практиков интересуются этой темой и применяют разнообразные методы анализа данных для

изучения правовых вопросов и проблем. Теперь, благодаря развитию междисциплинарного подхода, юристы могут активно использовать эти методы и инструменты в своей работе и исследованиях.

Междисциплинарный подход позволяет ученым из разных научных областей объединять свои знания и опыт для изучения сложных и многоаспектных проблем, которые могут возникнуть в юридической сфере. Это приводит к новым открытиям, которые повышают эффективность и справедливость правовой системы. Таким образом, юристы могут учиться у экономистов, социологов и других ученых, применяя количественные и качественные методы в своих исследованиях и практике. Это включает в себя использование статистического анализа, моделирования и прогнозирования, а также применение методов интервью, глубинных опросов и других качественных инструментов.

Важно отметить, что использование междисциплинарного подхода не означает отказа от традиционных юридических методов исследования, но скорее их дополнение и расширение новыми инструментами и подходами. Это позволяет юристам получить более полное и глубокое понимание правовых вопросов и проблем, что способствует созданию более грамотных и обоснованных решений и законодательных актов.

Междисциплинарный подход является мощным инструментом для юристов и исследователей, который может помочь им расширить свои знания и навыки, улучшить практику и повысить качество исследований. Это также способствует развитию сотрудничества между учеными из разных областей науки и практики, что может привести к новым открытиям и инновациям, способствующим развитию и совершенствованию правовой системы и правопорядка. Подход, включающий сотрудничество с социологами, экономистами, политологами и другими специалистами, может помочь юристам лучше понимать, как реальные люди взаимодействуют с правовой системой, и какие изменения могут сделать эту систему более отзывчивой на их потребности. Одним из примеров того, как юристы могут использовать социологические методы и инструменты, является проведение социологических опросов для выявления реальных потребностей и ожиданий населения от правовой системы. Опросы могут дать ценную информацию о том, какие законы и регулирование на самом деле работают, и какие аспекты правовой системы нуждаются в улучшении.

Я бы хотел рассказать о том, какие исследования мы проверили на арбитражных данных.

## Как юристы и социологи воспринимают одни и те же данные?

Для юриста арбитражные данные, как правило, являются источником информации о конкретных судебных делах, прецедентах и правовых позициях, которые необходимы для анализа и решения своих практических задач. Юристы, как правило, ищут информацию по конкретным вопросам, сравнивают практику разных судов и определяют наиболее вероятный исход дела на основе имеющихся данных.

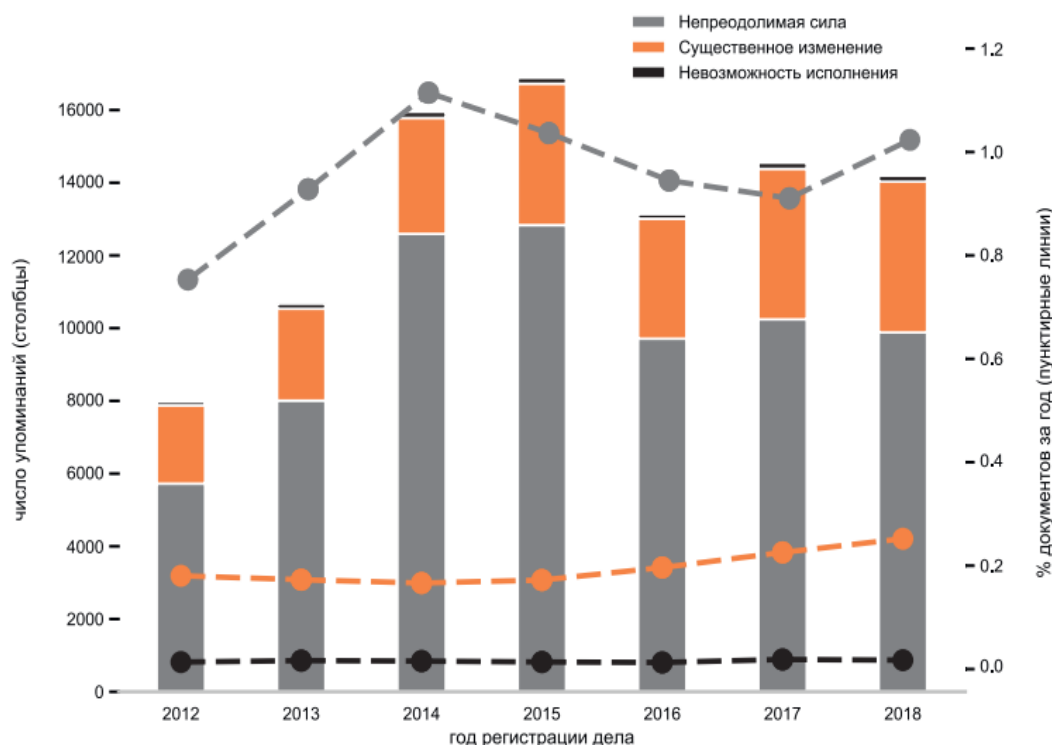
Социологи же, используя те же арбитражные данные, могут проводить более широкие исследования общественных и социальных аспектов правоприменения. Они анализируют статистику, выявляют тенденции и закономерности в поведении участников правоотношений, а также изучают влияние правовых решений на общество в целом. Такой подход позволяет получить более глубокое понимание того, как функционирует правовая система, и какие факторы влияют на ее эффективность.

Междисциплинарные исследования, основанные на арбитражных данных, могут привести к открытию новых знаний и пониманию важных аспектов правоприменения. Например, анализ данных может помочь выявить проблемы в законодательстве, которые приводят к большому количеству споров и судебных разбирательств, или выявить области, где отсутствует четкость и определенность правил, что создает условия для коррупции и злоупотреблений.

Таким образом, использование арбитражных данных в междисциплинарных исследованиях, проводимых в рамках сотрудничества юристов и социологов, позволяет получить ценную информацию о реальном функционировании правовой системы и ее воздействии на общество. Эти знания могут быть использованы для совершенствования законодательства и правоприменительной практики, а также для обучения будущих юристов, которые должны быть готовы эффективно работать в сложных и динамичных условиях современного мира.

Использовались эти понятия в различных отраслях экономики. Мы выявили, что в некоторых секторах, таких как транспорт, туризм и ресторанный бизнес, обращение к форс-мажору и существенным изменениям обстоятельств было значительно чаще, чем в других секторах, что логично, учитывая специфику их работы и наибольшую уязвимость к внешним воздействиям.

**Рисунок 1. Частота упоминаний терминов, связанных с невозможностью исполнения обязательств**



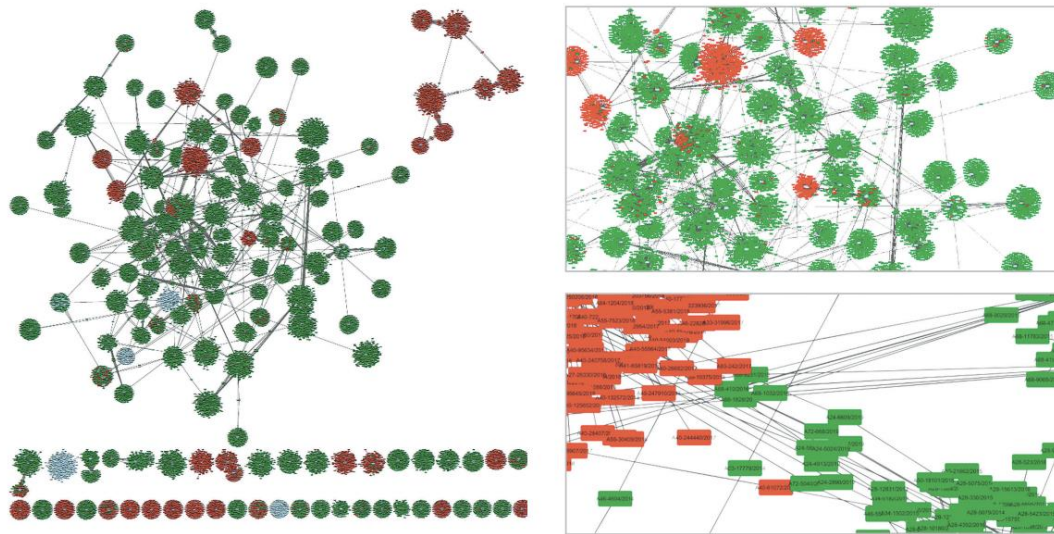
На столбчатой диаграмме показано абсолютное число упоминаний указанных терминов в год в документах, в которых таких упоминаний 2 и более. Это число отложено по оси Y слева. Подсчитывалось каждое упоминание в документе. Пунктирными линиями отложена доля (в процентах) числа документов, содержащих два и более таких упоминания, среди всех дел за указанный год в используемой выборке. Доля отложена по оси Y справа.

Таким образом, изучение большого количества текстов и анализ частотности использования определенных понятий позволяет получить представление о том, какие аспекты правоприменения являются наиболее актуальными и востребованными в определенный период времени, и какие проблемы возникают в связи с изменением обстановки.

Важным результатом такого исследования является то, что оно позволяет выявить пробелы в законодательстве и практике его применения, которые становятся особенно заметными в периоды кризиса и социально-экономических изменений. Эти данные могут быть использованы для корректировки нормативных актов и разработки новых правовых механизмов, а также для подготовки юристов и судей, которые должны быть готовы адекватно реагировать на сложные ситуации и принимать обоснованные решения.

Мы также проводили исследование чтобы выяснить как часто суды ссылаются на решения других судов:

Рисунок 1. Карта сети ссылок в арбитражных делах(с увеличением фрагментов)



Прим. На рисунке представлена карта ссылок одних дел на другие по номеру и отдельно увеличенный участок сети. Рисунок создан программой *cytoscape* с использованием силового алгоритма визуализации графов (англ. *prefuse force directed layout*)<sup>1</sup>. Красным цветом отображены административные дела, зеленым — гражданские дела, голубым — иные типы дел. На рисунке представлена лишь некоторая часть всех ссылок (показаны связи только на те дела, на которые ссылаются от 100 до 500 других дел, а также участники процесса не совпадают). Интерактивная карта, созданная в Институте проблем правоприменения доступна по адресу <http://precedent.enforce.spb.ru>.

В целом, изучение реальной правовой практики, а не только текстов нормативных актов, позволяет получить более полное и глубокое понимание правовой системы и ее функционирования в обществе. Это дает возможность выявить слабые места и потенциал для улучшений, что в свою очередь способствует повышению эффективности правоприменительной деятельности и укреплению веры граждан в правосудие.

# Анализ сложности НПА: на стыке права и компьютерной лингвистики

Александр Чаплинский

Правовые нормы выражаются через язык. От качества этого языка зависит возможность правильно применять нормы. Я предлагаю вашему вниманию наше исследование, посвященное этой теме.

Когда мы занялись этим вопросом то начали с гипотезы: «Язык нормативных актов становится все сложнее с каждым годом.» Мы это чувствовали, но хотели доказать. Сразу появился вопрос – существуют ли какие-то методики для измерения сложной текста? Да, таких методик много.

Индексы	Длина предложения (в словах)	Длина слова (в слогах)	Количество букв (или слогов)	Количество предложений	Количество слов	Количество слов с более чем 3 слогами	Количество сложных слов
Формула Flesch-Kincaid	ДА	ДА					
Индекс Колман-Лиау			ДА	ДА			
Автоматический индекс удобочитаемости (Automatic Readability Index)			ДА		ДА		
SMOG				ДА		ДА	
Формула Dale-Chall	ДА						ДА



Сервис оценки читабельности русскоязычных текстов (<http://ru.readability.io/>)  
© Иван Бегтин (с) 2014

Но у них есть ряд недостатков. Изначально эти методики разрабатывались для английского языка и даже их адаптации плохо работают с русским языком. В основном, метрики, которые используются, сводятся к длине предложений и сложности слов в тексте. Однако, эти методики не

учитывают множество других аспектов, которые могут влиять на сложность текста, особенно когда речь идет о юридических документах.

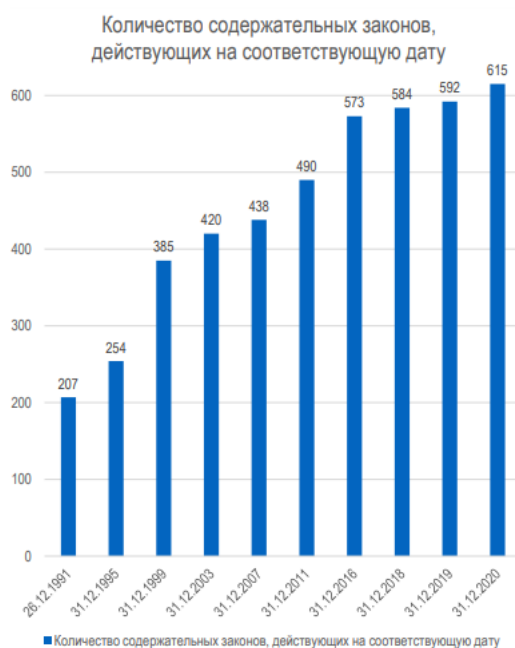
Например, они не учитывают структуру текста, наличие определений и ссылок на другие документы, а также логическую связность аргументов.

Поэтому мы решили разработать собственную методику для измерения сложности текстов правовых актов, которая будет учитывать специфику русского языка и особенности юридических текстов. Для этого мы использовали машинное обучение и анализ текстовых данных, чтобы выявить наиболее значимые характеристики, влияющие на сложность текста.

С чего мы начали?

## Логика исследования

### 1. Формирование корпусов законодательных актов для оценки удобочитаемости



Год	Среднее кол-во статей в законе	Среднее количество слов в законе	Среднее количество слов в статье закона
1991	51	4970	96
1995	43	5000	117
1999	39	5460	139
2003	41	6935	164
2007	42	8200	190
2011	42	9100	217
2016	42	11140	263
2018	42,7	12150	280
2019	42,5	12450	291
2020	42,2	12544	296

С каждым годом количество законов увеличивается. Сама структура текста становится больше, хотя количество статей и не меняется.

## 2. Лингвистический анализ корпуса законодательных актов Российской Федерации: установление «нормальных» значений параметров

На этом этапе с использованием средств компьютерной лингвистики мы проанализировали синтаксические параметры текстов по 9 метрикам:

Параметр	Средние значения								Л.Н. Толстой Анна Каренина	Современный русский литературный язык		
	1991	1995	1999	2003	2007	2011	2016	2020		А.В. Иванов Географ глобус пропил (фрагмент)	Д.А. Глуховский Текст	Е.Г. Водолазкин Авиатор
Доля глаголов от общего количества слов в тексте (%)	5,4 %	4,6 %	4,3 %	4,2 %	4,1 %	4 %	3,8 %	3,8%	18%	18%	19%	17%
Отношение количества зависимых слов от существительных к общему количеству существительных (не являющихся зависимыми) в тексте*	2,9	3,2	3,35	3,4	3,4	3,57	3,66	3,7	1,2	0,74	0,75	0,95
Количество причастных оборотов после определяемого слова (на одно предложение)	0,3	0,35	0,34	0,33	0,35	0,42	0,47	0,5	0,04	0,02	0,008	0,02
Количество слов в одном предложении	18	18,7	18,56	18,2	18	19,3	20,3	20,9	13,63	8,42	7,5	9
Расстояние между зависимыми словами (количество слов, отделяющее определяемое слово от зависимого от него)	2,35	2,3	2,4	2,4	2,4	2,46	2,47	2,5	2,14	1,92	2	1,96
Количество предиктивных ядер в предложениях	1,1	1,1	1,12	1,12	1,14	1,17	1,2	1,2	1,45	1,14	1,2	1,3
Количество слов в абзаце (количество слов)	21,3	22,76	23,3	23,3	23,8	27	29,7	30,9	37	19,1	21	32,3

Дальше нам нужно было определить как сами читатели воспринимают те или иные метрики, то есть насколько текст с высокими или низкими показателями по разным метрикам, является сложным для чтения.

## 3. Определение относительной значимости параметров и формулы расчета сводного индекса сложности

Мы отобрали фрагменты текстов, где по очереди был превышен каждый параметр и один простой текст. Проводили эксперимент на профессиональных юристах и студентах. Им надо было прочитать текст и оценить сложность, а также ответить на контрольные вопросы.

- ✓ эксперимент
- ✓ 10 фрагментов законодательных актов (в каждом 1 параметр превышен + «простой» фрагмент)
- ✓ 21 профессиональный юрист и 20 студентов юридической магистратуры МГУ и НИУ ВШЭ
- ✓ оценка удобочитаемости фрагментов и ответы на контрольные вопросы

### Рейтинг параметров

Параметр*		К-т значимости
Доля глаголов в оцениваемом тексте	ГЛ	18%
Количество слов в субстантивных именных словосочетаниях в каждом предложении оцениваемого текста	СИС	16%
Количество слов в абзацах в каждом абзаце оцениваемого текста	ПР	15%
Расстояние между зависимыми словами в оцениваемом тексте	РЗС	15%
Количество причастных оборотов, расположенных после определяемого слова, приходящихся на 1 предложение оцениваемого текста	ПО	14%
Количество слов в предложении в каждом предложении оцениваемого текста	СП	11%
Количество грамматических основ (предикативных основ, предикативных ядер) предложения (подлежащее – сказуемое или одно из них) в каждом предложении оцениваемого текста	ГО	11%

\* для параметра ГЛ действует правило «Чем меньше, тем хуже», для остальных параметров – «Чем больше, тем хуже»

Так мы рассчитали коэффициент значимости соответствующих параметров.

#### 4. Расчет индекса синтаксической сложности нормативных правовых актов

Из полученных данных мы смогли рассчитать ИСЗ. Он учитывает не только средние, но и «пиковые» показатели. Это важно, потому что в акте могут быть сложными отдельные статьи, из-за чего повышается уровень сложности всего документа.

- ✓ проводится для текста закона, структурированного на статьи
- ✓ учитывает средние значения и «пиковые значения» по отдельным статьям

$$\begin{aligned}
 \text{ИСЗ} = & \text{норм\_ГЛ} \times 0,18 + \text{норм\_СИС}_{\text{средн}} \times 0,08 + \text{норм\_СИС}_{\text{макс}} \times 0,08 + \\
 & + \text{норм\_ПР}_{\text{средн}} \times 0,075 + \text{норм\_ПР}_{\text{макс}} \times 0,075 + \text{норм\_РЗС}_{\text{средн}} \times 0,075 + \\
 & + \text{норм\_РЗС}_{\text{макс}} \times 0,075 + \text{норм\_ПО}_{\text{средн}} \times 0,07 + \text{норм\_ПО}_{\text{макс}} \times 0,07 + \\
 & + \text{норм\_СП}_{\text{средн}} \times 0,055 + \text{норм\_СП}_{\text{макс}} \times 0,055 + \text{норм\_ГО}_{\text{средн}} \times 0,055 + \\
 & + \text{норм\_ГО}_{\text{макс}} \times 0,055) \times 100
 \end{aligned}$$

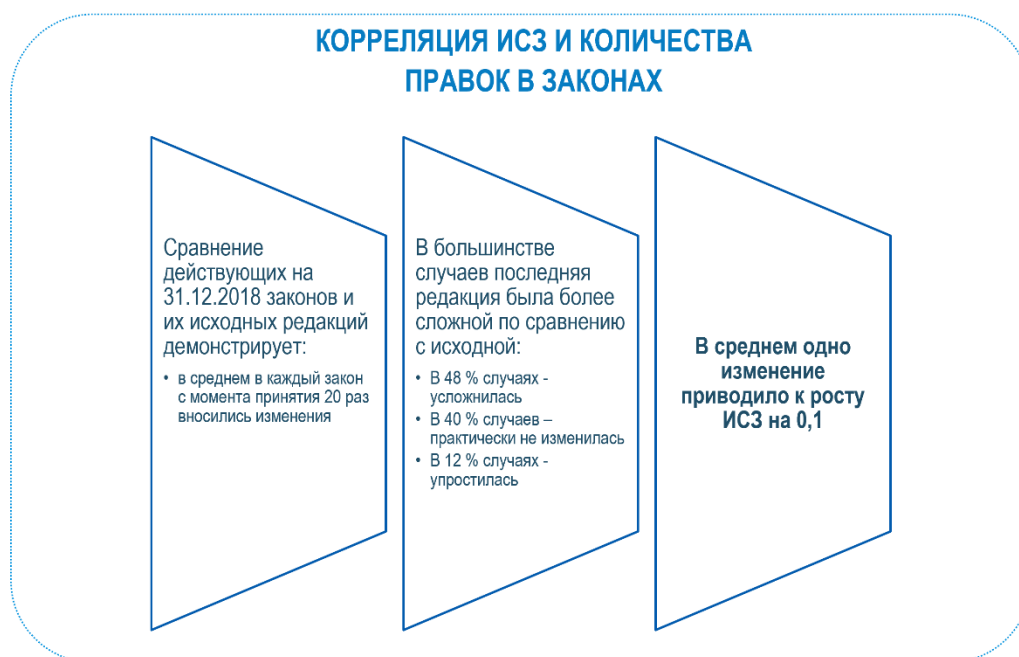
Важно отметить, что наш индекс синтаксической сложности имеет определенные ограничения. Например, он не учитывает длину закона или длину статьи закона, а также не различает синтаксические конструкции, построенные на перечислениях.

Мы разработали специальный сервис (программное средство) для оценки сложностей законодательных актов, что позволило нам

установить индекс синтаксической сложности НПА. Мы обнаружили, что некоторые действующие федеральные законы имеют более высокий индекс сложности, чем философские произведения (например, «Критика чистого разума» И. Канта).

Среди наиболее сложных законов можно найти многие законы, касающиеся особенностей правового регулирования в связи с принятием в состав России Республики Крым и города федерального значения Севастополь. Возможно, это было связано с высокой скоростью разработки таких актов, а также с необходимостью внесения точечных изменений («правовых заплаток») в другие законы. Если говорить о нормотворчестве в целом, то «правовые заплатки» обычно оказываются более сложными, чем тексты, написанные с нуля.

Мы также проанализировали сложность наиболее распространенных и активно применяемых законов, а также изучили динамику сложности законодательства в разные годы и по различным отраслям. На основе этого анализа можно сделать вывод о том, как влияют на сложность законов принятие поправок и новых актов. В некоторых случаях принятие нового акта взамен старого приводит к усложнению текста, но иногда бывает и наоборот - новый закон становится проще.



Помимо академической ценности исследования, мы также разработали методические рекомендации по повышению читаемости текстов. В нашей методичке представлены правила, которые объясняют, почему определенное правило важно, приводятся примеры негативных случаев и указывается, как исправить сложности в конкретных элементах текста.

**Методические рекомендации по повышению удобочитаемости  
текстов НПА: как писать простым языком?**

**Правило № 1:** Краткие и емкие термины

**Правило № 2:** Короткие и простые предложения

**Правило № 3:** Компактное расположение пары «подлежащее - сказуемое» и использование действительного залога

**Правило № 4:** Упрощение словосочетаний

**Правило № 5:** Меньше повторов

**Правило № 6:** Умеренное количество причастных оборотов

**Правило № 7:** Сбалансированный размер статей и других структурных единиц текста

**Правило № 8:** Новый нормативный акт вместо «лоскутного одеяла»

Мы разработали инструмент, который позволяет оценить сложность текста, и он размещен на [сайте Вышки](#). Инструмент предоставляет возможность вводить текст вручную, а также загружать файлы с текстами. После обработки текста инструмент автоматически рассчитывает синтаксические метрики и значение индекса сложности. В дополнение к этому, инструмент имеет функцию цветовой раскраски сложных мест текста. Вы можете настроить параметры выделения сложных мест, изменяя пороговое значение индекса сложности для выделения. Таким образом, пользователи могут увидеть, какие части текста являются наиболее сложными с точки зрения синтаксиса и структуры.

Это исследование мы создали в рамках проекта «Речевые практики российского общества». Мы отвечаем за юридический язык. Мы определенно рассматриваем возможность сделать наш инструмент более доступным и открытым для широкой аудитории. В настоящее время инструмент доступен на сайте нашего университета, и мы постоянно работаем над его улучшением и развитием.

Мы будем продолжать улучшать наш инструмент и искать возможности сделать его более открытым и доступным для широкого круга пользователей. Ваши предложения и отзывы, конечно же, важны для нас, и мы будем учитывать их в дальнейшем развитии инструмента.

# Ответы на вопросы

## Вопрос из зала

Екатерина, какие произошли изменения в обучении по сравнению с использованием старых методов?

### Екатерина Безызвестных

Изменения в обучении связаны с тем, что мы начинаем использовать данные для анализа и улучшения образовательных программ и подходов. Это действительно отличается от традиционного подхода, когда данные не использовались или использовались в меньшей степени. Однако стоит отметить, что эти изменения не означают полного отказа от старых методов, а скорее их дополнение новыми подходами, инструментами и метриками, которые позволяют нам лучше понимать результативность обучения и находить пути для его улучшения. Следует также учесть, что внедрение этих изменений может вызывать определенные трудности и сопротивление со стороны преподавателей и студентов, которые привыкли к старым подходам. Важно проводить работу по информированию и обучению всех участников образовательного процесса, объяснять причины и цели изменений и показывать преимущества новых подходов, чтобы смягчить этот переход и сделать его более плавным и эффективным.

## Вопрос из зала

Денис, как вы получаете доступ к данным?

### Денис Савельев

Эти исследования арбитражных данных мы получали от компании Право.ру под нераспространение для научных исследований, некоторые реестры покупали. Для проверки законодательства мы собирали данные с официальных сайтов гос. органов. К сожалению, открытых данных остается все меньше.

## Вопрос из зала

Александр, почему происходит осложнение текста и что мешает участникам законотворческого процесса писать понятнее?

### Александр Чаплинский

Да, конечно мы думали об этом и у нас появилось несколько гипотез. Первая гипотеза – психологическая. Людям кажется что написать понятный текст значит показать свою непрофессиональность.

Если что-то написано на обыденном языке, значит не качественно. Вторая гипотеза – текст усложняется на этапе правок, чтобы ликвидировать проблемы в уже действующих актах. Желание учесть все нюансы приводит к усложнению текста. А глобальная проблема – это порыв к избыточной точности. Это желание поместить в документ как можно больше деталей. По-хорошему наши гипотезы следовало бы проверить путем опроса.

### **Вопрос из зала**

Александр, скажите, пожалуйста, вы не думали сделать этот инструмент открытым? Чтобы можно было проверить текст по разным метрикам и даже внести правки в автоматическом режиме.

#### **Александр Чаплинский**

Этот инструмент находится в общем доступе на сайте Высшей школы экономики. Воспользоваться им может любой. Единственное, для общего пользования открыта проверка либо одного файла, либо возможность вписать текст от руки. Да, этот сервис не так популярен, как другие, и надо заняться его продвижением, что я сейчас и делаю.

### **Вопрос из зала**

Студенты часто спрашивают: есть ли смысл идти по пути воспитания законодателя в плане написания более понятных текстов, если скоро искусственный интеллект можно будет попросить систематизировать наше законодательство?

#### **Александр Чаплинский**

Конечно, в будущем искусственный интеллект и нейронные сети могут сыграть большую роль в написании и систематизации законов, но пока что они не могут полностью заменить человека, особенно в сложных юридических текстах. Поэтому на текущем этапе актуально продолжать обучение и воспитание законодателей в написании более понятных и доступных законов, а также использовать доступные инструменты для анализа и улучшения сложности текстов.

Да, стоит популяризировать то, чтобы люди писали проще, доступнее, и это будет полезно не только для нас, но и для общества в целом. Если бы больше людей обратило внимание на эту проблему и начало стремиться к написанию более понятных и доступных законов, это бы значительно улучшило качество законодательства и облегчило его применение и понимание для широкой аудитории.

**Денис Савельев**

Позвольте мне дополнить коллегу. Тот, кто ясно мыслит, тот ясно излагает. Иногда бывает трудно написать просто.

### **Вопрос из зала**

Как вы считаете, коллеги, могут ли сети-трансформеры как ChatGPT помочь нам в изучении языка массивов текстов? Например, в GPT есть огромный корпус нормативных актов. Раньше для решения специфических задач человек неделю сидел и программировал, анализировал акты, полностью погружался в тему. Сейчас это можно попросить выполнить GPT.

**Денис Савельев**

У генеративных сетей есть большая проблема – они выдумывают факты. Это очень опасно для применения. Инструменты надо тщательно подбирать.

С легкими задачами они лучше справляются. Системы, построенные на контекстах, находят больше информации. Но при этом точность будет меньше, и мы не сможем назвать количественные данные. Мы можем получить наиболее релевантный ответ на вопрос, но не знаем будет он придуман или построен на реальных данных. GPT – новый инструмент, который все изучают, но не стоит забывать о методах анализа текста и нужно выбирать наиболее подходящий.

### **Вопрос из зала**

Коллеги, мы говорим про сложность законов и про исследование языка алгоритмическими методами, что подразумевает разбиение текста на части для его анализа. Были ли у вас попытки, идеи, когда изучалась сама логика юридического текста? Есть языки, построенные на формальной логике. Гипотетически, с использованием нейросетей можно построить инструмент, который будет переводить тексты законов на язык формальной логики и потом искать пробелы будет легче.

Просто у меня есть один воодушевляющий пример: мой друг экспериментирует с трансформерами в плане перевода текстов на формальные языки. Он говорит, что при некоторых доработках методик, можно было бы попробовать перевести большой массив текста. Раньше такой возможности вообще не было. Говорили, что для того чтобы переписать Гражданский кодекс на машиночитаемый язык нужно посадить целый НИИ математиков. От этой идеи отказались, а сейчас она возвращается.

У нас в компании есть такой проект. Это сложный проект. Я бы сказал, что это совсем не простой путь. Для получения промышленно применимых результатов потребуется еще много времени.

**Александр Чаплинский**

На мой взгляд, такая методология для законодательных, нормативных актов применима с трудом. Есть правовые документы, для которых это может оказаться более перспективно, например, судебные решения. Потому что судебные решения должны быть более структурированными: описание обстоятельств, аргументы с обеих сторон, их оценка судом и так далее. Хотя даже для судебных решений это тяжело сделать.