

Реферат

Выпускная квалификационная работа 133 с., 50 рис., 37 табл., 19 источников, 3 прил.

Ключевые слова: БАЗА ДАННЫХ, OLAP, DATA MINING, БОЛЬШИЕ ДАННЫЕ, BUSINESS INTELLIGENCE, СУДЫ.

Объектом исследования является задача нахождения скрытых закономерностей, а предметом исследования – решения по судебным делам.

Цель работы – анализ записей, полученных в результате обработки документов и решений судов Томской области с помощью методов и средств технологии Business Intelligence.

В процессе исследования проводился анализ предметной области судопроизводства Томской области, кроме того был выполнен обзор существующих аналогов.

В работе приведены результаты проектирования и реализации базы данных, а также проектирование и разработка инструментов для извлечения параметров, необходимых для проведения анализа, и наполнения базы данных судопроизводства. Приведены результаты OLAP анализа данных судебных решений Томской области.

В дальнейшем, разработанные инструменты могут позволить заинтересованным людям самостоятельно извлекать данные и проводить их анализ. Анализ можно расширить путем увеличения объема анализируемых решений и проводить анализ не только по судам Томской области, а по всем судам Российской Федерации.

Оглавление

Введение	10
1 Анализ предметной области судопроизводства	12
1.1 Актуальность анализа данных судопроизводства	12
1.2 Анализ структуры данных судебных решений	13
1.3 Обзор аналогов	15
1.3.1 «CaseLook»	15
1.3.2 «РосПравосудие»	17
1.4 Постановка задачи	18
2 Проектирование инструментов для обработки данных	20
2.1 Выбор инструментов разработки	20
2.2 Проектирование базы данных судопроизводства	21
2.3 Проектирование инструмента для извлечения атрибутов судебного решения	23
2.4 Проектирование инструмента для наполнения базы данных	31
3 Разработка инструментов для обработки данных	34
3.1 Разработка инструмента для извлечения атрибутов судебного дела ..	34
3.2 Разработка инструмента для наполнения базы данных	35
4 Анализ данных с использованием технологии OLAP	37
5 Финансовый менеджмент, ресурсоэффективность и ресурсосбережения	73
6 Социальная ответственность	109
Заключение	120
Список публикаций	122
Список использованных источников	123
Приложение А	125
Приложение Б	126
Приложение В	127

Введение

В наше время происходит интенсивное накопление огромных объёмов данных разного типа в различных предметных областях измеряемые в петабайтах, это в свою очередь даёт возможность решать задачи получения новых фактов, зависимостей и скрытых корреляций, а также позволяет решать некоторые аналитические задачи, такие как прогнозирование, проверка статистических гипотез, расчёт агрегатных показателей и т.д.

Объектом исследования является задача нахождения скрытых закономерностей, а предметом исследования – решения по судебным делам.

Цель работы – анализ записей, полученных в результате обработки документов и решений судов РФ с помощью методов и средств технологии Business Intelligence.

Анализ судопроизводства довольно актуален, поскольку результаты анализа будут интересны как большим организациям, таким как правоохранительные органы и правозащитные организации, так и отдельным лицам, таким как независимые юристы и люди, находящиеся под следствием. Для больших организаций будет более интересен глубокий анализ для поиска скрытых зависимостей и отклонений, а отдельным лицам анализ по конкретным судьям и адвокатам, который может помочь предсказать какое решение может быть вынесено определенным судьей или выбрать наиболее эффективного адвоката.

Значимость данной работы состоит в том, что разработанные инструменты позволят заинтересованным людям самостоятельно извлекать и анализировать данные.

В первом разделе данной работы проведен анализ предметной области, обоснована актуальность исследования и изложены результаты обзора аналогов.

Во втором разделе представлены проектирование и реализация базы данных, а также приведены результаты проектирования инструментов для

извлечения атрибутов из текстов судебных дел и наполнения базы данных судопроизводства.

В третьем разделе представлены результаты разработки инструментов для извлечения атрибутов из текстов судебных дел и наполнения базы данных судопроизводства.

В четвертом разделе представлены результаты OLAP анализа данных, извлеченных из решений по судебным делам.

В пятом разделе рассмотрены вопросы финансового менеджмента, результаты расчетов и анализ ресурсоэффективности и ресурсосбережения.

В шестом разделе описаны основные требования социальной ответственности.

Извлечение данных выполнено в среде PyCharm, язык программирования – Python. Запись данных в БД выполнена в среде Visual Studio 2013, язык программирования – C#. Для создания БД использовалась среда Microsoft SQL Server 2014 Management Studio. OLAP и Data Mining анализ выполнен в среде Visual Studio 2013, с использованием Excel 2013. Пояснительная записка оформлена в Microsoft Word 2013.

1 Анализ предметной области судопроизводства

1.1 Актуальность анализа данных судопроизводства

Впервые термин «большие данные» использовал Клиффорд Линч в выпуске журнала Nature от 3 сентября 2008 года, с темой «Как могут повлиять на будущее науки технологии, открывающие возможности работы с большими объёмами данных?». После этого направление обработки больших данных начало стремительно развиваться. Так, в 2010 году уже начали появляться первые программные продукты и решения для обработки больших данных.

1 июля 2010 года вступил в силу Федеральный закон «Об обеспечении доступа к информации о деятельности судов в Российской Федерации» [1], который предусматривает размещение решений по судебным делам на сайтах судов. Благодаря данному закону, появилась возможность открытого доступа к судебным решениям, и соответственно возможность выполнить анализ дел судопроизводства.

Таким образом, имеются не только инструменты необходимые для анализа, но и данные судопроизводства находятся в открытом доступе. Уже есть несколько проектов, занимающихся анализом судебных дел, однако, эти проекты либо не предоставляют возможности анализа вообще, либо набор функций довольно ограничен и доступ к ним платный.

Анализ судопроизводства довольно актуален, поскольку результаты анализа будут интересны как большим организациям, таким как правоохранительные органы и правозащитные организации, так и отдельным лицам, таким как независимые юристы и люди, находящиеся под следствием. Поскольку позволят не только собрать статистическую информацию, но и увидеть более глубокий анализ по отдельным судьям, адвокатам и другим атрибутам судебных дел для определения различных скрытых взаимосвязей между атрибутами дела и для поиска отклонений. Результат анализа может

помочь предсказать какое решение может быть вынесено определенным судьей или помочь выбрать наиболее эффективного адвоката.

Задачей данной работы является анализ решений по судебным делам, с помощью методов и средств Business Intelligence, для выявления скрытых закономерностей.

1.2 Анализ структуры данных судебных решений

Поскольку дела судопроизводства появились в открытом доступе сравнительно недавно, структура официальных веб-сайтов судов, на которых размещаются судебные дела, претерпевала изменения. Так, на некоторое время на сайтах судов появились данные о движении дела, однако довольно быстро эти данные были убраны. Поэтому пришлось ориентироваться исключительно на содержимое судебных актов, это связано с крайней разобщённостью судов в отношении публикаций движения дел, а также с отсутствием таковой информации на некоторых сайтах судов вообще.

На сайтах судов дела выложены после некоторой обработки, которая выделила некоторые атрибуты: имя судьи, дата суда, вид дела, подвид дела, инстанция, регион, город. Однако, это лишь часть необходимых атрибутов. Поэтому, был проведен анализ текстов судебных дел для определения списка атрибутов, которые необходимо и возможно извлечь. Основная сложность заключается в том, что тексты судебных дел довольно не структурированы и некоторые данные, например, имена некоторых ответчиков, обезличены.

Текст судебного дела можно разделить на четыре части:

1. В вводной части решения суда должны быть указаны дата и место принятия решения суда, наименование суда, принявшего решение, состав суда, секретарь судебного заседания, стороны, другие лица, участвующие в деле, их представители, предмет спора или заявленное требование.

2. Описательная часть решения суда должна содержать указание на требование истца, возражения ответчика и объяснения других лиц, участвующих в деле.

3. В мотивировочной части решения суда должны быть указаны обстоятельства дела, установленные судом; доказательства, на которых основаны выводы суда об этих обстоятельствах; доводы, по которым суд отвергает те или иные доказательства; законы, которыми руководствовался суд.

4. Резолютивная часть решения суда должна содержать выводы суда об удовлетворении иска либо об отказе в удовлетворении иска полностью или в части, указание на распределение судебных расходов, срок и порядок обжалования решения суда. Резолютивная часть решения суда, принятого мировым судьей, также должна содержать указание на срок и порядок подачи заявления о составлении мотивированного решения суда [2].

Таким образом, необходимые атрибуты судебного дела в основном содержатся в основной и резолютивной частях дела.

В результате анализа текстов судебных дел и сайтов судов Томской области, было выявлено следующее. По типу судебные дела делятся на:

- административные;
- гражданские;
- уголовные;
- по материалам.

В зависимости от типа судебного дела можно выделить типы судебных решений:

- постановления;
- решения;
- определения;
- приговоры.

Каждое судебное решение имеет следующие общие атрибуты:

- номер дела;

- город;
- регион;
- название суда;
- судья (ФИО, пол);
- дата суда;
- нормативный акт (номер статьи, название кодекса);
- дата вступления в силу;
- решение по делу;
- физические лица, фигурирующие в делах.

Кроме основных атрибутов судебное решение может иметь дополнительные атрибуты, в зависимости от типа дела и вынесенного решения:

- наказание;
- размер наказания;
- мера наказания;
- ответчик;
- истец;
- ФИО адвоката;
- юридические лица, фигурирующие в делах.

У одного дела может быть несколько решений, если в деле фигурирует несколько ответчиков. В свою очередь одно решение может содержать несколько наказаний, в случае вынесения нескольких видов наказаний (арест, взыскание и т.д.). Кроме того, у каждого решения есть свой список статей, на основании которых было вынесено решение.

1.3 Обзор аналогов

1.3.1 «CaseLook»

Сервис «CaseLook» позиционируется как универсальный инструмент для поиска и анализа судебной практики. «CaseLook» - это коммерческий проект, который предусматривает повременную оплату начиная от месяца до года [3].

Сервис «CaseLook» имеет следующие элементы для фильтрации судебных решений из имеющейся в данном сервисе базы данных:

Судебный акт:

- текст;
- тип документа;
- дата принятия;
- дата поступления;
- норма права.

Судья и суд:

- суд;
- судья;
- тип суда;
- нижестоящий суд;
- субъект РФ.

Дело:

- вид спора;
- категория спора;
- номер дела;
- исход дела;
- участник;
- представитель.

В данное время аналитические инструменты в сервисе «CaseLook» находятся в экспериментальном режиме и не представлены в полном объеме, хотя заявлены на сайте и оглашаются во время переговоров со службой поддержки.

Основные свойства сервиса:

- Огромная база судебных решений.
- Широкие поисковые возможности документов.
- Анализ документов.
- Постоянное обновление судебных актов.
- Оценка вероятности исхода дела (экспериментальный режим).

Положительные стороны:

- Удобный интерфейс пользователя.
- Предоставление тестового режима на две недели для ознакомления.
- Служба поддержки клиентов.

Недостатки:

- Высокая цена использования сервиса.
- Закрытая база данных.
- Ограниченный набор задач.
- Некоторые инструменты работают в тестовом режиме.

1.3.2 «РосПравосудие»

РосПравосудие - проект, созданный в 2012 году для исследования общедоступной судебной практики.

РосПравосудие обеспечивает возможность изучения судебной практики с помощью анализа текстов решений, публикуемых судами на основании Федерального закона «Об обеспечении доступа к информации о деятельности судов в Российской Федерации» от 22 декабря 2008 г. за № 262-ФЗ. Тексты решений получаются с официальных сайтов судов общей юрисдикции, арбитражных и мировых судов и отражают только те решения, которые опубликованы судами за последние несколько лет [4].

Внесение изменений в тексты документов и принятие решений об их публикации является исключительной компетенцией работников суда.

Навигация на сайте РосПравосудие происходит при помощи фильтров, последовательно сужающих перечень документов.

В качестве фильтра может использоваться все, что угодно - регион, суд, категория дела, интервал времени, участник процесса или фраза.

Проект РосПравосудие позиционирует себя как общественный проект призванный сделать судебную систему прозрачной.

Финансирование проекта осуществляется на добровольное пожертвование людей.

Основные свойства сервиса:

- Огромная база судебных решений.
- Широкие поисковые возможности документов.
- Анализ документов.
- Постоянное обновление судебных актов.
- Содержание на сайте некоторой статистики по судебным решениям.

Положительные стороны:

- Удобный интерфейс пользователя.
- Бесплатный доступ к сервису.
- Очень широкие возможности по фильтрации судебных решений.

Недостатки:

- Отсутствие аналитических инструментов в режиме реального времени.
- Закрытая база данных.
- Медленное развитие проекта из-за отсутствия постоянной финансовой поддержки.

1.4 Постановка задачи

В результате анализа предметной области, включающий обзор аналогов, было выявлено, что базы данных по судопроизводству нет в свободном доступе. Поэтому появилась необходимость самостоятельно

разработать БД, извлечь параметры судебного дела и наполнить БД. На рисунке 1 представлена общая схема подготовки данных для анализа, которая демонстрирует этапы, через которые необходимо пройти для проведения анализа данных судопроизводства.



Рисунок 1 – Общая схема подготовки данных для анализа

Для того, чтобы выполнить анализ данных судопроизводства, необходимо выполнить следующие задачи:

1. Выбор инструментов проектирования, разработки и проведения анализа.
2. Проектирование и реализация базы данных.
3. Проектирование и разработка инструментов для извлечения параметров и наполнения базы данных.
4. Проведение OLAP анализа данных.

2 Проектирование инструментов для обработки данных

2.1 Выбор инструментов разработки

Для извлечения параметров судебных дел использовался язык программирования Python в среде разработки Pycharm. Основными достоинствами выбранной среды и языка является:

- удобочитаемый синтаксис;
- прост в обучении;
- высокоуровневый язык;
- объектно-ориентированность;
- обширный набор подключаемых модулей.

Для создания БД использовалась среда Microsoft SQL Server 2014 Management Studio, за счет следующих достоинств:

- удобный графический интерфейс;
- богатые возможности;
- высокая производительность;
- надежность.

В качестве среды разработки приложения для наполнения БД была использована Visual Studio 2013. Выбор данного программного обеспечения обусловлен следующими предоставленными возможностями:

- Работа с объектами баз данных в обозревателе объектов SQL Server.
- Определение таблиц в новом конструкторе таблиц.
- Разработка и тестирование приложений баз данных в SQL Server Express LocalDB.
- Запись асинхронного кода простым и интуитивно понятным способом.
- Получение сведений о вызывающем объекте, которые помогают с трассировкой и отладкой.

В качестве языка программирования использован язык C#, так как он обладает следующими преимуществами:

- объектно-ориентированность;
- ориентация на безопасность кода;
- многофункциональность;
- относительная простота.

Для построения OLAP–кубов и проведения OLAP–анализ использовалась среда Visual Studio 2013, поскольку она имеет следующие достоинства:

- средства Business Intelligence;
- работа с базами данных SQL Server;
- удобный графический интерфейс.

2.2 Проектирование базы данных судопроизводства

На основе проведенного анализа структуры судебных решений была спроектирована схема базы данных.

База данных состоит из 25 сущностей со своими наборами атрибутов:

- Регионы (идентификатор региона, название региона);
- Города (идентификатор города, идентификатор региона, название города);
- Формы юридических лиц (идентификатор формы юридического лица, форма юридического лица);
- Юридические лица (идентификатор юридического лица, имя юридического лица, идентификатор формы юридического лица);
- Юридические лица дела (идентификатор юридического лица, идентификатор дела);
- Физические лица (идентификатор физического лица, имя физического лица);

- Физические лица дела (идентификатор физического лица, идентификатор дела);
- Истцы (идентификатор истца, тип истца, имя истца);
- Суды (идентификатор суда, название суда, идентификатор города);
- Виды дел (идентификатор вида дел, название вида дел);
- Подвиды дел (идентификатор подвида дел, название подвида дел);
- Дела (идентификатор дела, номер дела, идентификатор суда, идентификатор судьи, идентификатор адвоката, категория, дата суда, дата вступления в силу, идентификатор истца, идентификатор вида дела, идентификатор подвида дела, идентификатор инстанции);
 - Судьи (идентификатор судьи, ФИО, пол судьи);
 - Решения (идентификатор решения, идентификатор дела, идентификатор ответчика, идентификатор типа решения);
 - Тип решения (идентификатор типа решения, тип решения);
 - Инстанции (идентификатор инстанции, название инстанции);
 - Наказания (идентификатор наказания, идентификатор решения, идентификатор типа наказания, размер наказания, идентификатор типа меры наказания);
 - Статьи решения (идентификатор решения, идентификатор статьи);
 - Адвокаты (идентификатор адвоката, ФИО);
 - Типы наказаний (идентификатор типа наказания, название типа наказания);
 - Типы мер наказаний (идентификатор типа меры наказания, название типа меры наказания);
 - Кодексы (идентификатор кодекса, название кодекса);

- Статьи (идентификатор статьи, статья, идентификатор кодекса, номер статьи, идентификатор категории статьи);
- Категории статей (идентификатор категории статей, название категории статей).

У каждой сущности есть свой уникальный идентификатор, который позволяет однозначно определить элементы сущностей, а также помогает предотвратить дублирование данных. За исключением трех сущностей (Статьи решения, Юридические лица дел, Физические лица дел), которые имеют составной ключ, состоящий из двух соответствующих идентификаторов.

Кроме того, практически все сущности связаны с помощью бинарных связей, в данном случае используется связь один ко многим. Таким образом, Решение может принадлежать только одному Делу, но у Дела может быть множество Решений. Таким же образом связаны и остальные сущности: Юридические лица и Формы юридических лиц, Города и Регионы, Дела и Истцы, Суды и Города, Дела и Суды, Дела и Виды дел, Дела и Подвиды дел, Дела и Инстанции, Дела и Судьи, Дела и Адвокаты, Решения и Тип решения, Решения и Ответчики, Наказания и Решения, Наказания и Типы наказаний, Наказания и Типы мер наказаний, Статьи и Кодексы, Статьи и Категории статей. Исключение составляют сущности: Решения и Статьи, Дела и Юридические лица, Дела и Физические лица. В этих трех случаях между сущностями используется связь многие ко многим, сформированная за счет соответствующих промежуточных таблиц.

База данных создана в среде Microsoft SQL Server 2014 Management Studio в соответствии с данными проектирования. В приложении А представлена физическая схема БД.

2.3 Проектирование инструмента для извлечения атрибутов судебного решения

Извлечение данных представляет собой сложный процесс получения текстовых данных из одного или разных источников и приведения его к унифицированному виду. Обычно объектами парсинга являются веб-сайты, но также могут быть xml и pdf документы, и любые другие текстовые файлы и документы, хранящие информацию в текстовом виде.

В данной работе в качестве объектов для парсинга выступают xml документы. Xml файлы суда представляют собой набор нескольких параметров в xml тегах (имя судьи, дата суда, вид дела, подвид дела, инстанция, регион, город) и текст судебного дела.

Для получения параметров из текстов решений был разработан специальный инструмент. В котором применяются следующие методы и алгоритмы:

- Алгоритм шинглов — алгоритм, разработанный для поиска копий и дубликатов рассматриваемого текста в веб-документе. Шинглы — выделенные из текста подпоследовательности слов, идущих друг за другом. Выборка происходит внахлест, то есть каждый новый шингл сдвигается на одно слово [5].

- Расстояние Левенштейна между двумя строками — это минимальное количество операций вставки одного символа, удаления одного символа и замены одного символа на другой, необходимых для превращения одной строки в другую [6].

- Регулярные выражения — формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов. По сути это строка-образец («шаблон», «маска»), состоящая из символов и метасимволов и задающая правило поиска [7].

Для каждого атрибута судебного дела была написана собственная функция, с соответствующими регулярными выражениями и алгоритмом поиска.

На рисунке 2 представлен алгоритм извлечения атрибутов из текстов судебных дел. Часть функций (извлечение номера дела, ФИО адвоката, истца, физических лиц, юридических лиц) не зависят от результатов остальных функций. Однако некоторым функциям требуются данные из других функций. Так, для извлечения пола судьи требуется имя судьи, для решений требуется список ответчиков, для наказаний список ответчиков и решений, для статей список ответчиков и решений.

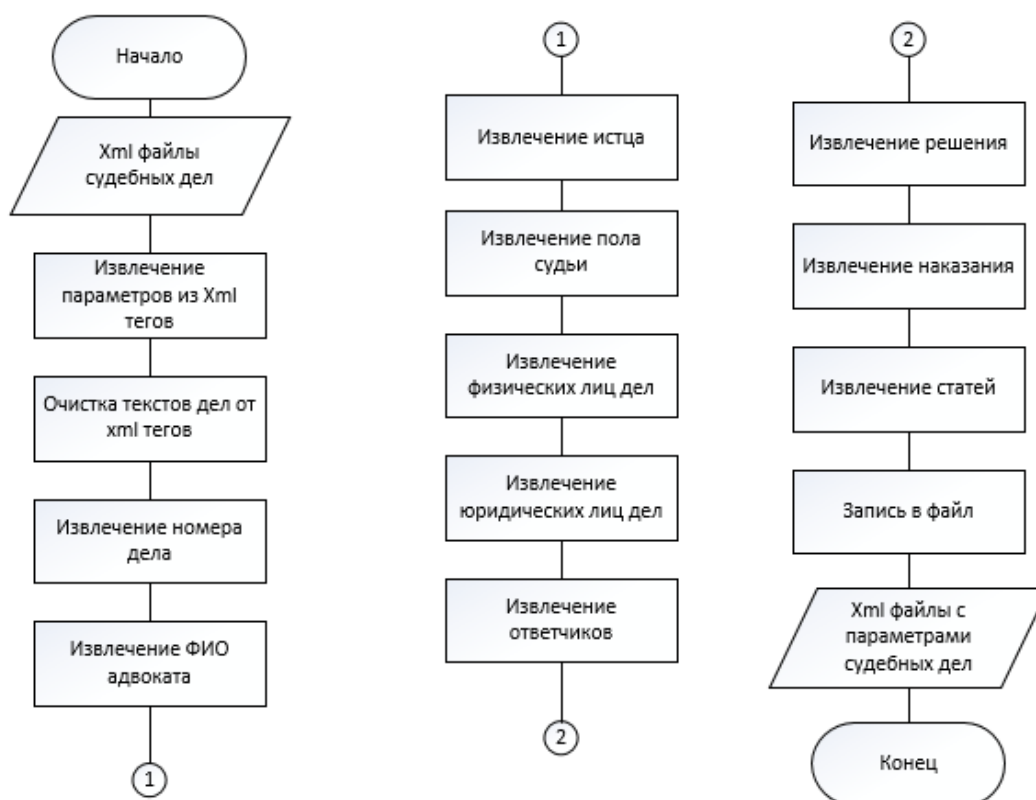


Рисунок 2 – Алгоритм извлечения атрибутов из текстов судебных решений

Извлечение параметров из xhtml тегов состоит в извлечении текста, заключенного в определенных xhtml тегах.

Выделение из текста дела резолютивной части, осуществляется с использованием регулярных выражений. Так, в тексте выполняется поиск заданных регулярных выражений (перед резолютивной частью ставятся следующие слова: постановление, решение и т.д.) при нахождении требуемого выражения выполняется обрезка текста от найденного выражения и до конца текста.

Выделение из резолютивной части дела фрагмента, относящейся к ответчику, осуществляется с использованием метода шинглов и расстояния Левенштейна. Поскольку имя ответчика может находиться в разных падежах в списке ответчиков и в тексте решения, текст разбивается на шинглы (с длиной равной количеству слов в имени ответчика) и используя расстояние Левенштейна сравниваются шинглы решения и шингл имени ответчика.

Алгоритм для извлечения физических лиц находит в тексте все выражения, удовлетворяющие заданному набору регулярных выражений. После чего, используя расстояние Левенштейна, из списка удаляются все дубликаты. Для поиска дубликатов используется расстояние Левенштейна, поскольку в тексте дела одни и те же лица могут упоминаться в разных склонениях. Такой же алгоритм используется при поиске юридических лиц (рисунок 3).

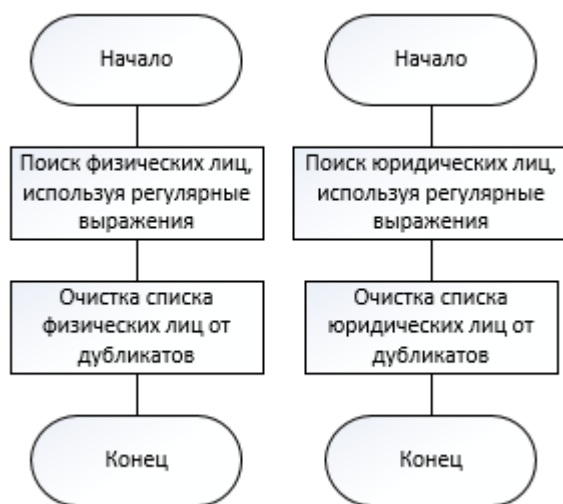


Рисунок 3 – Алгоритмы для извлечения физических и юридических лиц

Алгоритм для извлечения номера дела находит в тексте первое выражение, удовлетворяющее заданному набору регулярных выражений и отсекает ненужную более часть выражения, которая описывает слова, стоящие перед искомым номером дела. В качестве результата берется первое найденное выражение поскольку в деле может быть лишь один номер дела. Такой же алгоритм используется при поиске истца и адвоката. Алгоритм для

извлечения пола судьи отличается тем, что поиск регулярных выражений происходит в ФИО судьи (рисунок 4).

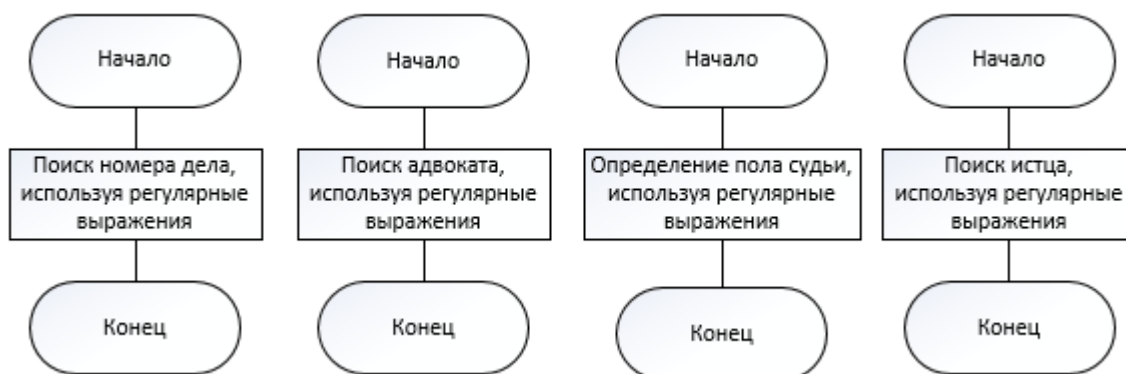


Рисунок 4 – Алгоритмы для извлечения номера дела, истца, адвоката и пола судьи

Алгоритм для извлечения ответчиков сперва выделяет часть текста с резолютивной частью. После этого идет по списку физических лиц и выделяет из решения часть текста, относящуюся к физическому лицу. Затем ищет в тексте заданные регулярные выражения, если выражение найдено, то в список ответчиков добавляется имя текущего физического лица. Если искомые выражения не найдены, то алгоритм переходит к следующему физическому лицу. Если не было найдено ответчиков среди физических лиц, то по такой же схеме, как для физических лиц, выполняется поиск ответчиков среди юридических лиц (рисунок 5).

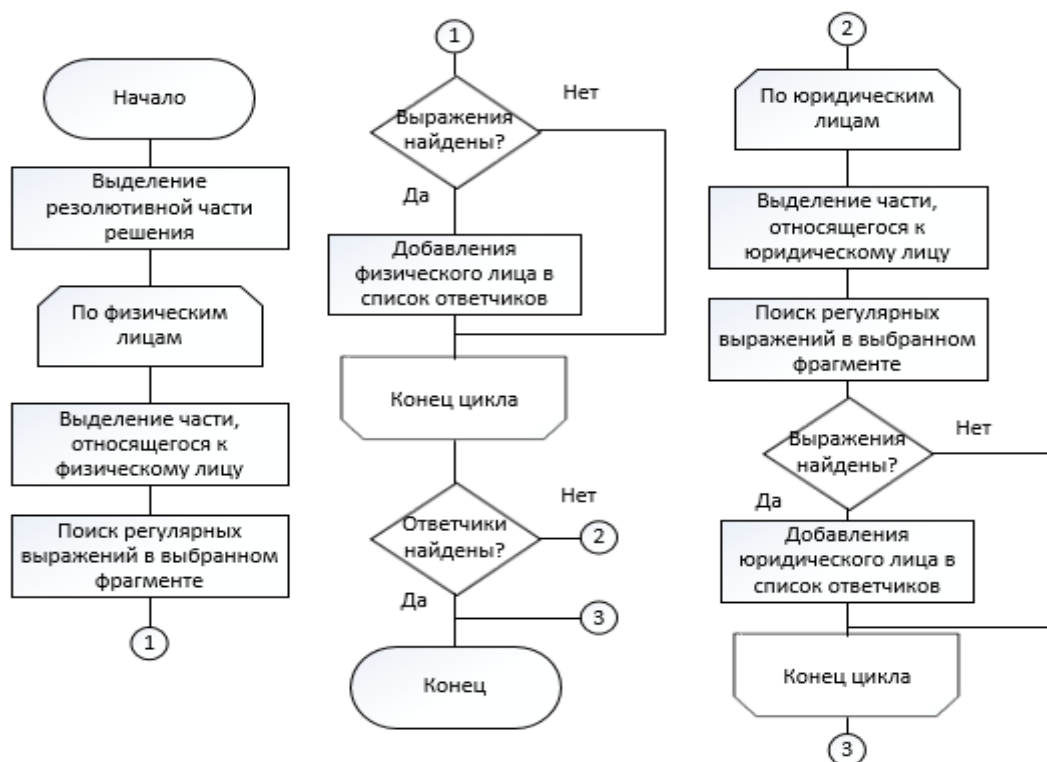


Рисунок 5 – Алгоритм для извлечения ответчиков

Алгоритм для извлечения решений сперва выделяет часть текста с резолютивной частью. Если есть ответчики, то алгоритм идет по списку ответчиков и выделяет из решения часть текста, относящуюся к ответчику. Затем ищет в тексте заданные регулярные выражения, если выражение найдено, то в список решений добавляется решение для текущего ответчика. Если искомые выражения не найдены, то алгоритм ведет поиск во всем решении по делу. Если же ответчиков нет, то поиск регулярных выражений ведется во всем тексте решения и результат записывается в список решений (рисунок 6).

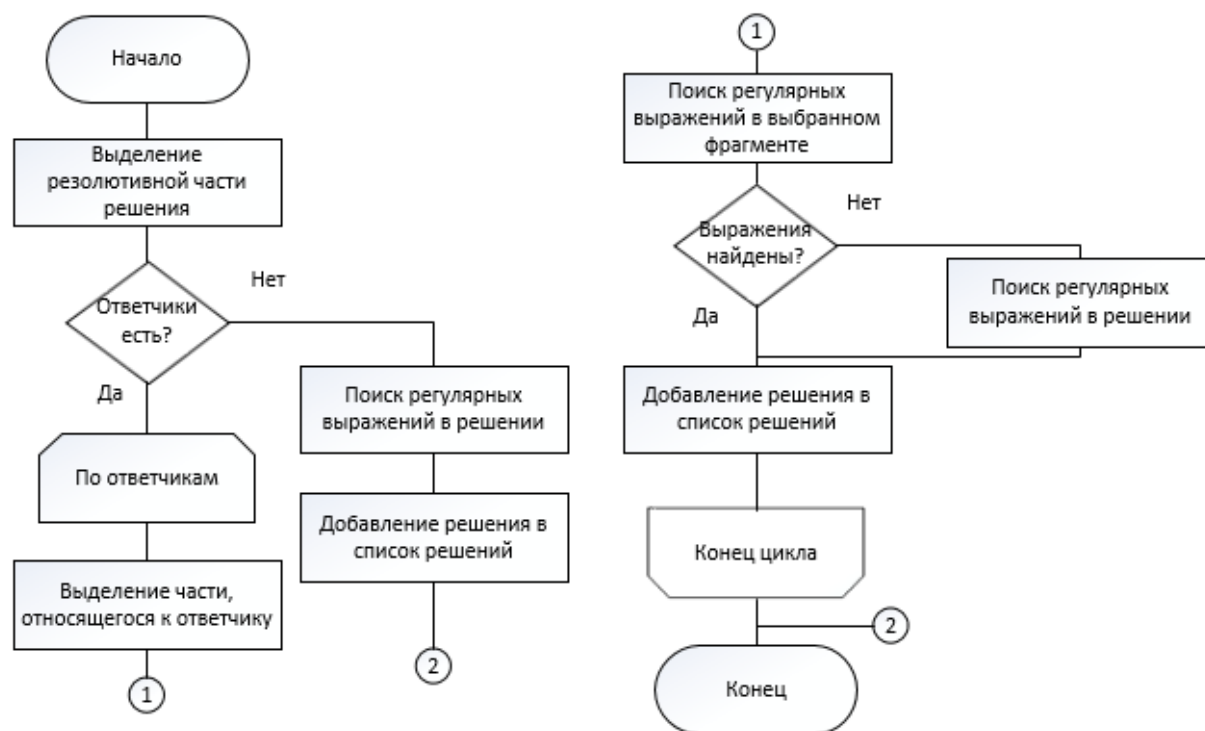


Рисунок 6 – Алгоритм для извлечения решений

Алгоритм для извлечения статей сперва выделяет часть текста с резолютивной частью. Если есть ответчики, то алгоритм идет по списку ответчиков и выделяет из решения часть текста, относящуюся к ответчику. Затем ищет в тексте регулярные выражения для номеров статей, если выражения найдены, то начинается поиск названия кодексов поблизости от номера статьи, с использованием регулярных выражений. Если название кодекса найдено, то в список статей добавляется номер статьи и название кодекса для текущего ответчика. Если кодекс не найден, то алгоритм переходит к следующему ответчику. Если в выделенной части решения с ответчиком не были найдены номера статей, то поиск производится во всем решении. Если номера были найдены, то далее идет поиск названия кодекса как уже было описано ранее. Если номеров статей нет во всем решении, то алгоритм завершает выполнение. Если же ответчиков нет, то поиск регулярных выражений проводится в тексте решения. И если номеров статей нет, то алгоритм завершает выполнение (рисунок 7).

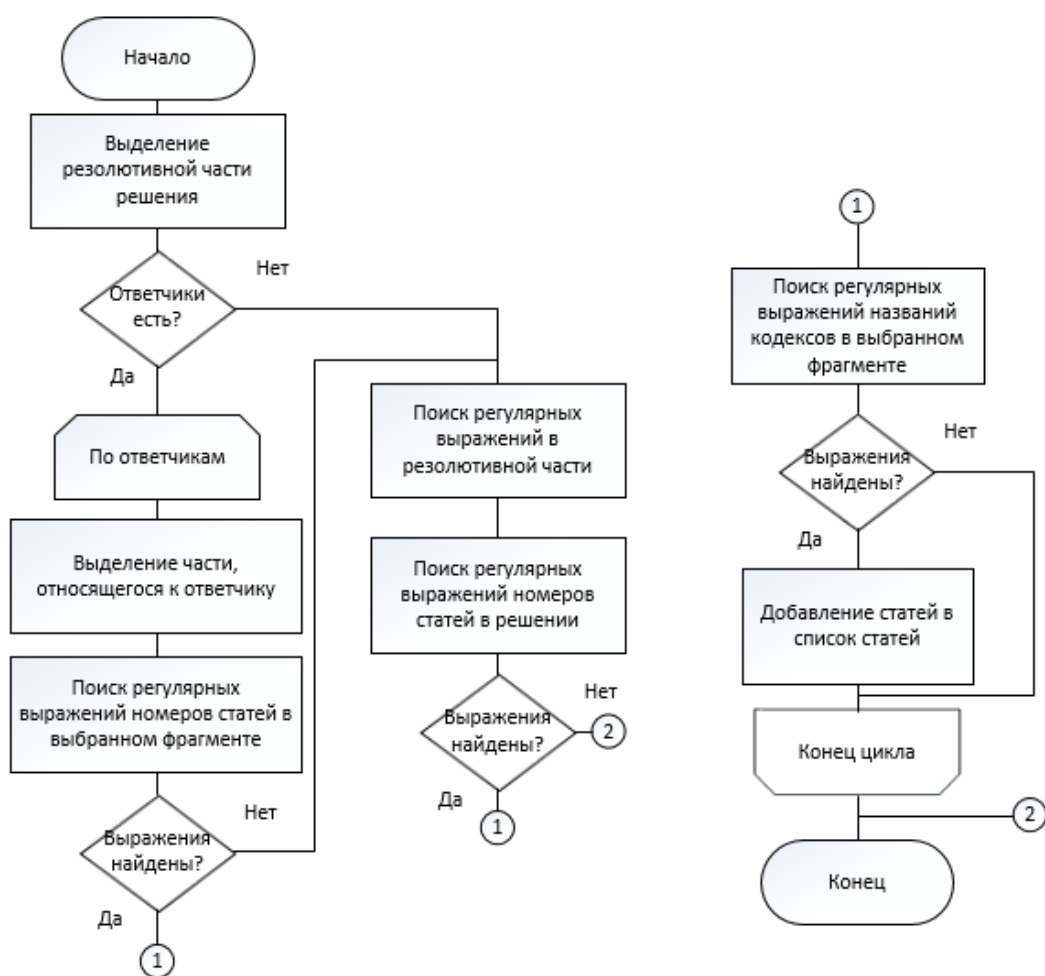


Рисунок 7 – Алгоритм для извлечения статей

Алгоритм для извлечения наказаний сперва выделяет часть текста с резолютивной частью. Если есть решения, то алгоритм идет по списку и выделяет из решения часть текста, относящуюся к текущему решению. Затем ищет в тексте заданные регулярные выражения наказаний, если выражение найдено, то в список наказаний добавляется наказание для текущего решения. Если искомые выражения не найдены, то алгоритм переходит к следующему и поиск ведется во всем решении по делу. Если же решений нет, то алгоритм завершает работу (рисунок 8).

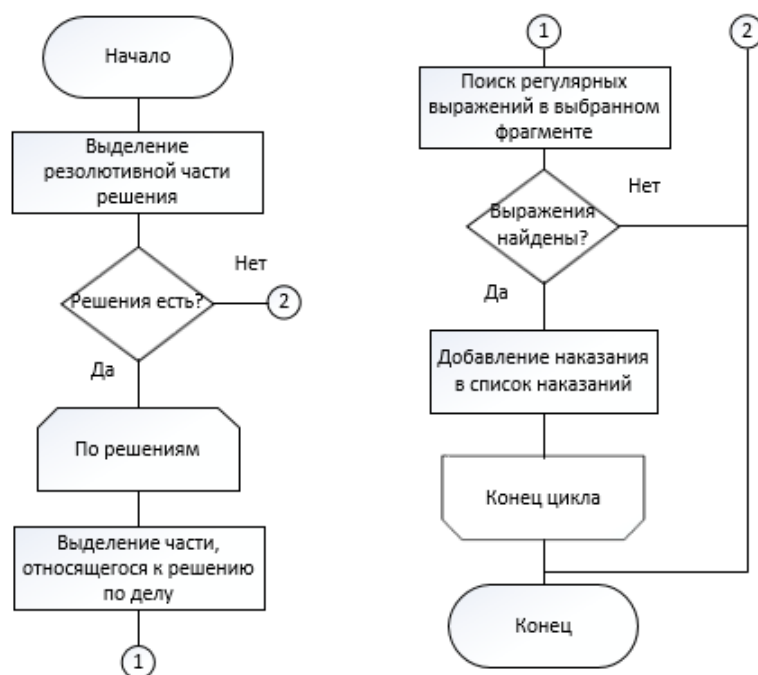


Рисунок 8 – Алгоритм для извлечения наказаний

Остальные атрибуты (имя судьи, вид и подвид дела, инстанция, дата суда, город и регион), находятся в xml тегах и очень просто извлекаются.

Извлеченные параметры записываются в xml файлы, каждый атрибут записывается на отдельную строку. Поскольку, атрибутов одного типа может быть несколько, то в файл записывается строка следующей структуры: «Название атрибута: Атрибут 1; ...; Атрибут N». Кроме того, у атрибута может быть несколько параметров (решение = имя ответчика + тип решения). То есть, «Атрибут 1» = «Параметр 1; ...; Параметр N».

2.4 Проектирование инструмента для наполнения базы данных

Наполнение БД довольно трудоемкое занятие, поскольку необходимо учитывать множество факторов. Так для каждой сущности необходимо выполнять проверку на дубликаты. Проверка выполняется с использованием расстояния Левенштейна, причем для атрибутов с именами нужно сравнивать не имя полностью, а отдельно фамилию и имя отчество. Поскольку в делах могут встречаться однофамильцы, поэтому в то время как фамилия может немного отличаться, за счет разных падежей, то имя отчество должны

проверяться на полное совпадение. Также имя и отчество может быть записано в виде инициалов или полное имя и отчество.

Также необходимо учитывать связи между различными сущностями. Так, например, нельзя сперва создать решение по судебному если еще этого дела еще нет, поскольку сущность «Решение» ссылается на идентификатор сущности «Дело». Кроме того, необходимо учитывать тип атрибута, так как БД требует точного их соблюдения.

У каждой сущности есть свой набор атрибутивной информации, включающий первичный и внешние ключи. У каждого экземпляра сущности есть свой уникальный ключ (он может состоять более чем из одного поля), который однозначно идентифицирует экземпляр. Внешние ключи, позволяют однозначно определить на какой экземпляр сущности 2 ссылается экземпляр сущности 1 [MS SQL Server 2012. Создание баз данных и разработка программ. Руководство для начинающих и профессионалов Бондарь Александр Григорьевич БХВ-Петербург, 2013 - Всего страниц: 608].

Упрощенный алгоритм записи в БД атрибутов судебного дела:

1. Первыми заполняются экземпляры сущностей, у которых нет внешних ключей (Адвокаты, Виды дел, Подвиды дел, Инстанции, Города, Судьи, Истцы).
2. Запись города, используя уникальный ключ региона.
3. Запись суда, используя уникальный ключ города.
4. Запись дела со всеми его атрибутами и внешними ссылками.
5. Записываются экземпляры сущность Юридические лица и одновременно заполняется промежуточная таблица юридических лиц, участвующих в деле.
6. Записываются экземпляры сущность Физические лица и одновременно заполняется промежуточная таблица физических лиц, участвующих в деле.

7. Затем, идет запись решений, вынесенных по делу. Причем, для каждого решения сперва записываются в соответствующие сущности ответчиков и типов вынесенного решения.

8. Далее для каждого решения записываются наказания, вынесенные по данному решению. Одновременно для каждого наказания записывается тип наказаний и тип меры наказания в соответствующие сущности.

9. Также для каждого решения записываются статьи, на которые оно ссылается. Перед записью статьи записывается название кодекса этой статьи в сущность, хранящую кодексы. После записи статьи в сущность добавляется запись в промежуточную таблицу статей решений.

Данный алгоритм выполняется для каждого файла с параметрами судебного дела.

3 Разработка инструментов для обработки данных

3.1 Разработка инструмента для извлечения атрибутов судебного дела

Инструмент для извлечения атрибутов судебного дела реализован на языке Python с использованием его различных расширений, позволяющих работать над извлечением текстовой информации.

Основной проблемой при реализации инструмента было, то что у судебных дел нет единой структуры и имена ответчиков обезличены. Поэтому, для некоторых параметров приходилось составлять довольно большой список регулярных выражений. Причем, эти списки приходилось постоянно дополнять, поскольку решения по разным делам довольно сильно отличались. Таким образом, реализация инструмента проводилась итеративно.

В результате удалось получить следующие данные:

- номер дела;
- область;
- город;
- суд;
- вид дела;
- тип дела;
- инстанция;
- ФИО Судьи;
- пол судьи;
- дата вступления в силу;
- дата суда;
- вынесенное решение;
- название нормативного акта;
- статья;

- наказание;
- размер наказания;
- адвокат;
- ответчик;
- истец;
- юридические лица;
- физические лица.

Извлеченные атрибуты записываются в xml файлы (рисунок 9).

```

judge:Прохорова Людмила Петровна
region:Томская область
city:Томск
Advokat:Лодыгиной Е.В.
persony:Ясюкевич Е.И.;Истомина А.Г.;Лодыгиной Е.В.;Туров
С.А.;Председательствующий Л.П.;Прохорова Л.П.
Stat:Туров С.А.|ч. 1 ст. 111|Уголовный кодекс Российской
Федерации
court:Бакчарский районный суд (Томская область)
vid_dokumenta:приговор
podsud:Туров С.А.
urlica:ОГУ|Томский территориальный фонд ОМС;ОГУ|Томский
территориальный фонд обязательного медицинского страхования
number:№ 1-20/2011
etapd:Первая инстанция
vidpr:Уголовное
Nakaz:Туров С.А.|Взыскание|84951|рубли;Туров С.А.|Арест|2|годы
Reshenie:Туров С.А.|Виновен
Istec:Истомина А.Г.
Pol:ж
DateSuda:13.04.2011
DateVstVSily:2011-02-18

```

Рисунок 9 – Xml файл, хранящий извлеченные атрибуты

3.2 Разработка инструмента для наполнения базы данных

На основе результатов проведенного этапа проектирования был реализован инструмент для заполнения базы данных судопроизводства на языке C# в среде Visual Studio 2013. На входе инструмент получает набор Xml файлов (рисунок 6).

Основной проблемой разработки инструмента было правильно реализовать взаимосвязи между различными сущностями и предусмотреть

все возможные варианты значений атрибутов. Так, имена могут записываться по-разному: фамилия имя отчество, фамилия инициалы имени и отчества, инициалы имени и отчества фамилия. Поэтому сперва имя разбивалось на отдельные слова, затем определялось где фамилия, а где имя и отчество. После этого проводится сравнение фамилий расстоянием Левенштейна и отдельно сравнение имени и отчества на совпадение.

В результате работы инструмента была получена база данных наполненная атрибутами судебных дел. Все атрибуты были записаны в соответствующие поля сущностей.

4 Анализ данных с использованием технологии OLAP

Business Intelligence - процесс извлечения многоаспектной информации и превращение её в знания для эффективного управления бизнесом, осуществляемый конечными пользователями с помощью специальных технологий, методов и средств [8]. Одной из технологий BI является OLAP.

OLAP - это оперативный анализ данных. 12 определяющих принципов OLAP сформулировал в 1993 г. Е. Ф. Кодд - автор реляционных БД. Позже его определение было переработано в так называемый тест FASMI, требующий, чтобы OLAP-приложение предоставляло возможности быстрого анализа разделяемой многомерной информации [9].

OLAP предоставляет удобные быстродействующие средства доступа, просмотра и анализа информации. Пользователь получает естественную, интуитивно понятную модель данных, организуя их в виде многомерных кубов. Осями многомерной системы координат служат основные атрибуты анализируемого бизнес-процесса. В качестве одного из измерений используется время. На пересечениях осей (измерений) находятся данные, количественно характеризующие процесс - меры. Пользователь, анализирующий информацию, может сделать срез по разным направлениям, получать сводные (например, по годам) или, наоборот, детальные (по неделям) сведения и осуществлять прочие манипуляции, которые ему придут в голову в процессе анализа [10].

Поскольку из текстов судебных решений было выделено достаточно много атрибутов, то можно проводить анализ с учетом разных атрибутов. Например, можно рассмотреть сезонный фактор, попытаться установить взаимосвязь между судьей и адвокатом. Мы решили рассмотреть, в качестве примера, влияние фактора пола судьи на различные параметры, такие как, наказание, решение и другие. Таким образом, с помощью OLAP анализ планируется решить следующие задачи:

- Сбор статистической информации.
- Проверить влияет ли пол судьи на принятие решений.
- Проверить влияет ли пол судьи на принятие решений по отдельным статьям и главам Уголовного Кодекса РФ.
- Проверить влияет ли пол судьи на вынесенное наказание.
- Проверить имеют ли судьи склонность к какому-либо решению или виду судебных дел.
- Посмотреть статистику по адвокатам.

Алгоритм построения куба:

1. Создаём проект в среде разработки Visual Studio 2013.
2. Выбираем существующую базу данных в качестве источника данных, для дальнейшей с ней работы.
3. Настраиваем представление источника данных и измерения. Кроме основных измерений (таблиц БД) было создано дополнительное измерение времени, включающий различные размерности (год, месяц, день и т.д.).
4. Создаём куб, при этом указывая какие измерения и группы мер будут в нём фигурировать. По структуре куб аналогичен спроектированной базе данных (Приложение Б).
5. После построения куба, необходимо проверить все ли связи между сущностями были интерпретированы верно. Кроме того, необходимо указать связи многие ко многим, поскольку автоматически они не распознаются.

Даже трехмерный куб сложно отобразить на экране компьютера так, чтобы были видны значения интересующих мер, не говоря уже о кубах с большей размерностью. Для визуализации данных, хранящихся в кубе, применяются, как правило, привычные двумерные, т.е. табличные представления, имеющие сложные иерархические заголовки строк и столбцов [11].

Используя полученный OLAP-куб получены следующие результаты:

1. По 89774 делам принято 90399 решений.
2. Из 108 судей 62 – женского пола, а 46 – мужского.
3. 7 судов: Советский, Тегульдетский, Бакчарский и Кожевниковский районные суды, Томский гарнизонный военный суд, Северский городской суд, Томский областной суд.
4. Распределение решений по видам дел (рисунок 10). Из среза видно, что больше 50% решений относятся к гражданским делам (вертикальная ось – процент от общего количества решений).

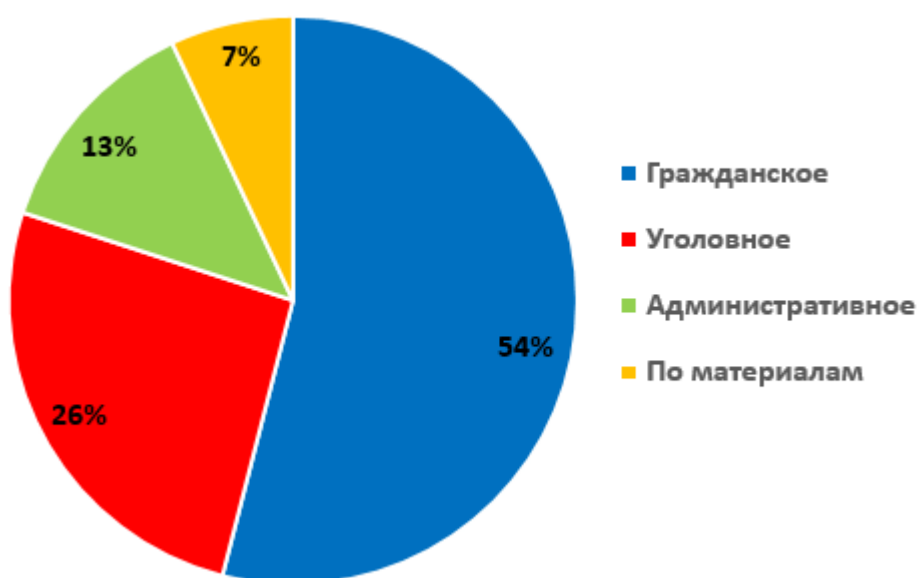


Рисунок 10 – Распределение решений по видам дел

5. Распределение решений по типу вынесенного решения (рисунок 11). Из среза видно, что преобладают три типа решения: Иск не удовлетворен, Иск удовлетворен и Виновен (вертикальная ось – процент от общего количества решений).

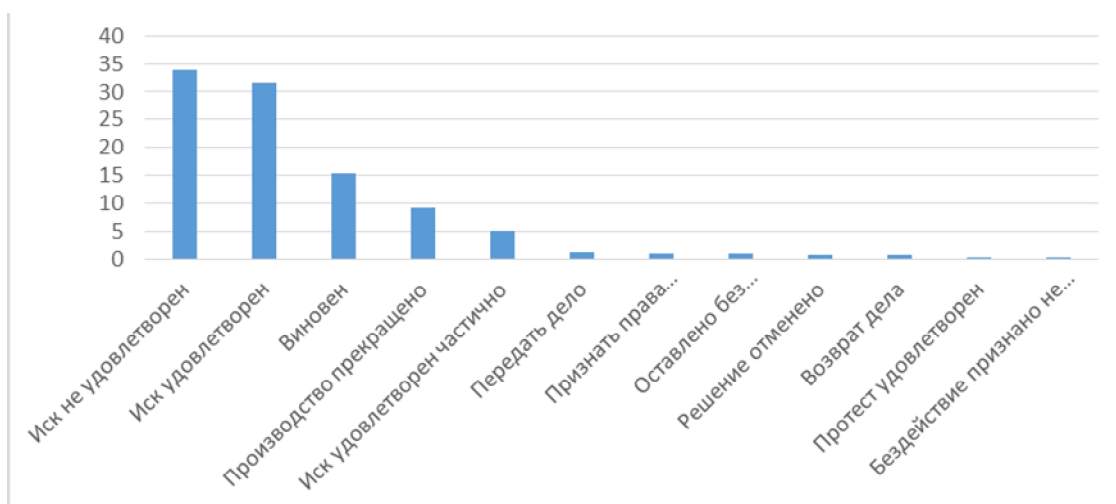


Рисунок 11 – Распределение решений по типу вынесенного решения

6. Распределение типов решений в зависимости от пола судьи (рисунок 12, 13). Данные этого среза были отфильтрованы по значению количества решений (типы решений со значением меньше 50 отбрасываются). Из среза видно, что судьи женского пола вынесли больший процент решений, чем судьи мужского пола, по следующим типам дел: Иск удовлетворен, Иск удовлетворен частично, Оставлено без рассмотрения и Протест удовлетворен. А судьи мужского пола вынесли больший процент решений по типам: Иск не удовлетворен, Виновен, Решение отменено и Бездействие признано не законным. По остальным типам процент решений между судьями разных полов не сильно отличается (вертикальная ось – процент от общего количества решений). На рисунке 13 типы были отсортированы по величине разницы между решениями судей.

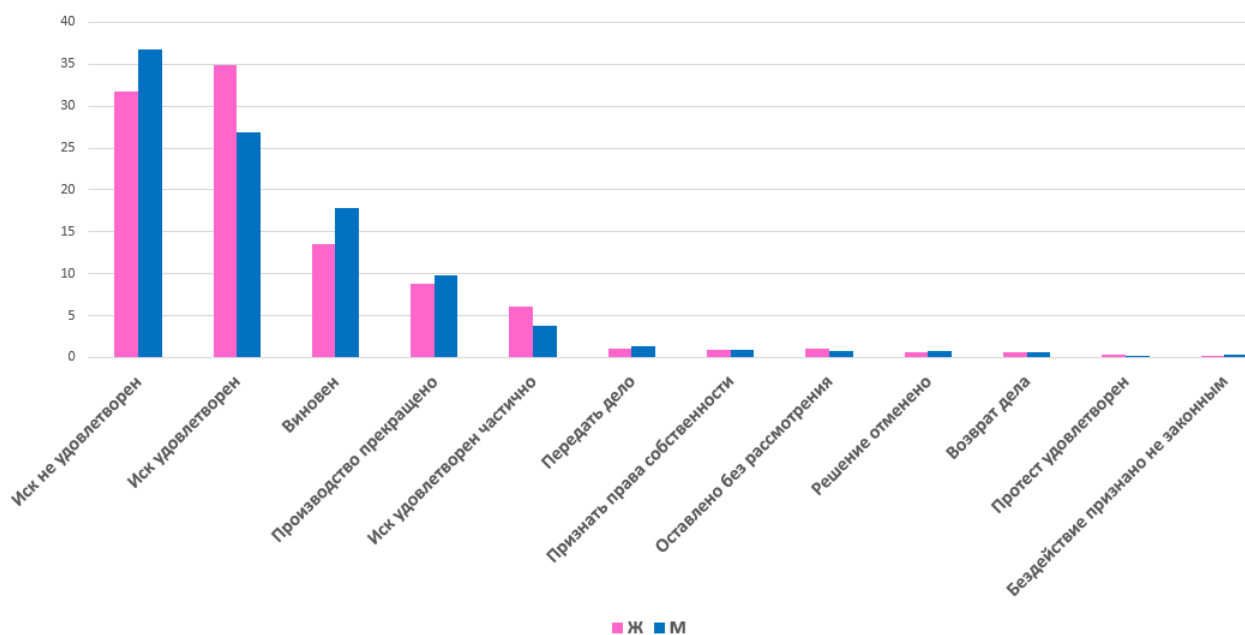


Рисунок 12 – Распределение типов решений в зависимости от пола судьи, гистограмма с группировкой

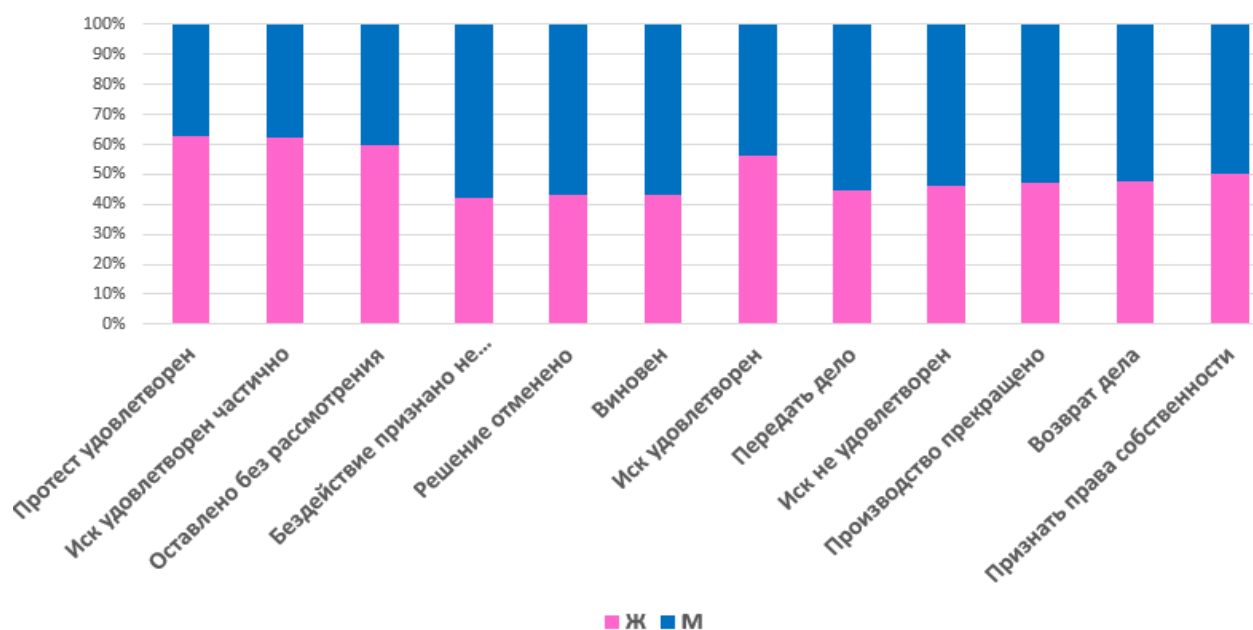


Рисунок 13 – Распределение типов решений в зависимости от пола судьи, нормированная гистограмма с накоплением

7. Распределение наказаний (рисунок 14). Из среза видно, что большая часть наказаний приходится на: Взыскание и Арест (вертикальная ось — процент от общего количества наказаний). А меньше всего предупреждений (всего 7 решений).

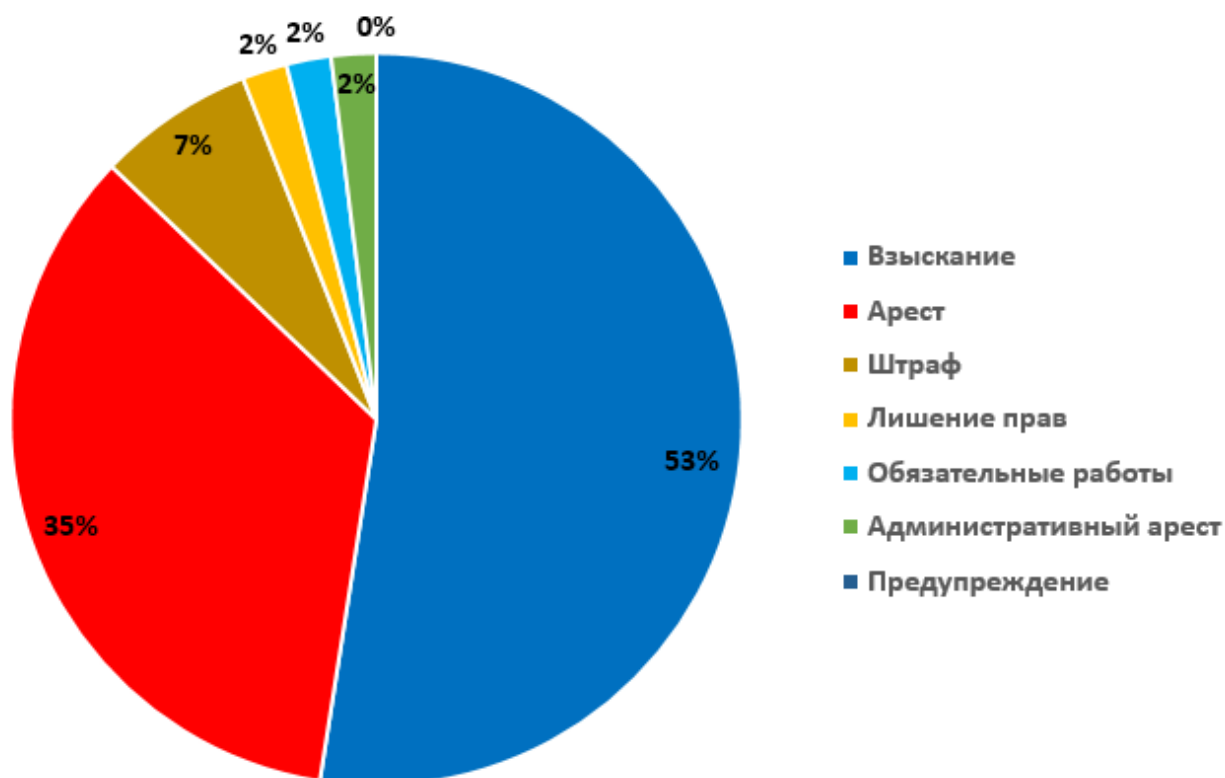


Рисунок 14 – Распределение наказаний

8. Распределение наказаний в зависимости от пола судьи (рисунок 15, 16). Данные этого среза были отфильтрованы по значению количества решений (наказания со значением количества решений меньше 50 отбрасываются). Из среза видно, что судьи женского пола вынесли больший процент наказаний, чем судьи мужского пола, по следующим типам наказаний: Взыскание. А судьи мужского пола вынесли больший процент по типам наказаний: Арест, Лишение прав и Обязательные работы. По остальным типам процент наказаний между судьями разных полов не сильно отличается (вертикальная ось – процент от общего количества наказаний). На рисунке 16 типы наказаний были отсортированы по величине разницы между решениями судей.

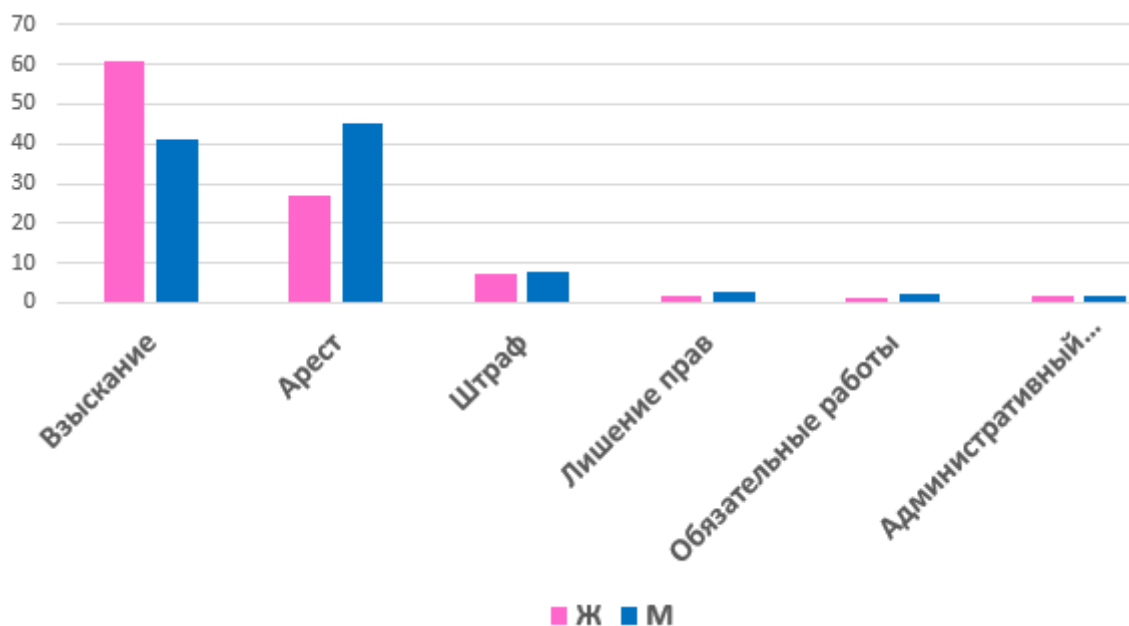


Рисунок 15 – Распределение наказаний в зависимости от пола судьи, гистограмма с группировкой

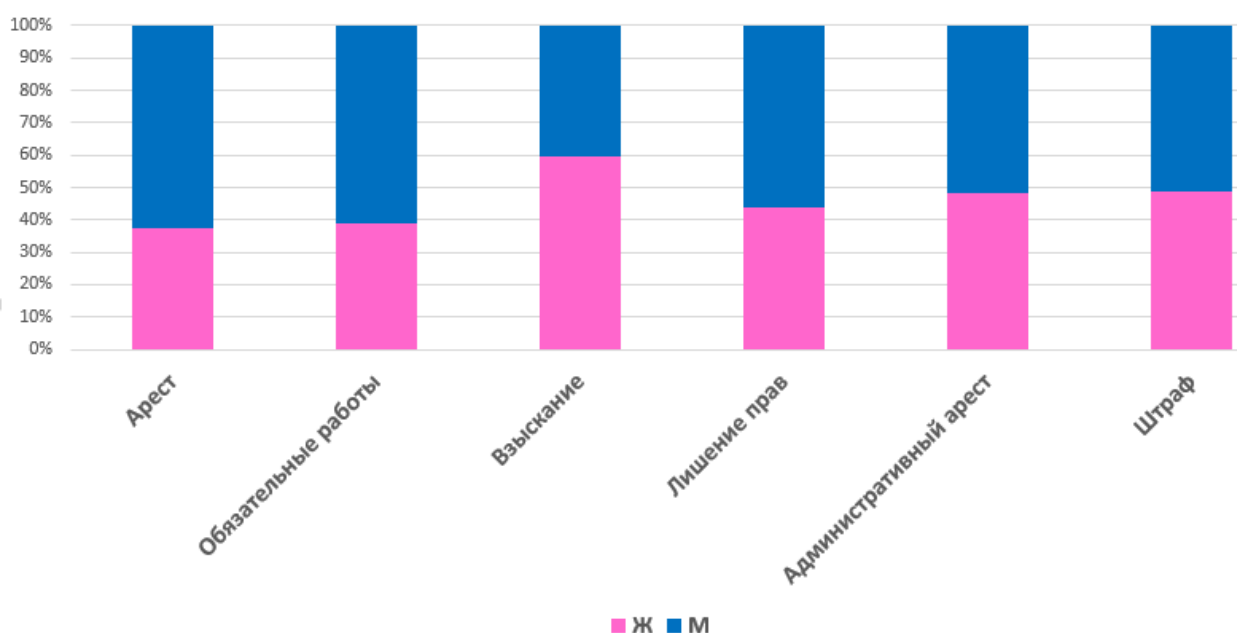


Рисунок 16 – Распределение наказаний в зависимости от пола судьи, нормированная гистограмма с накоплением

9. Распределение количества решений в зависимости от номеров статей Уголовного кодекса РФ (рисунок 17). Данные этого среза были отфильтрованы по значению количества решений (статьи со значением количества решений меньше 30 отбрасываются), вертикальная ось диаграммы – процент от общего количества решений по уголовным делам, которые ссылаются на УК РФ. Из среза видно, что наиболее

распространенными являются преступления, предусмотренные следующими статьями УК РФ: 158 (Кража), 228 (Незаконные приобретение, хранение, перевозка, изготовление, переработка наркотических средств, психотропных веществ или их аналогов, а также незаконные приобретение, хранение, перевозка растений, содержащих наркотические средства или психотропные вещества, либо их частей, содержащих наркотические средства или психотропные вещества), 161 (Грабеж) и 111 (Умышленное причинение тяжкого вреда здоровью) [12].

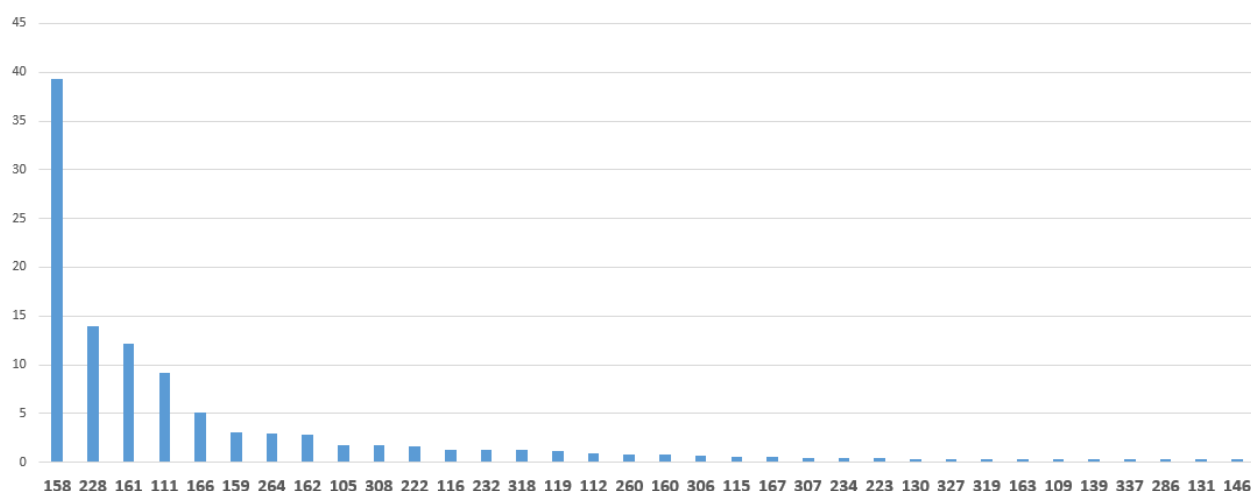


Рисунок 17 – Распределение количества решений в зависимости от номеров статей Уголовного кодекса РФ

10. Распределение типов решений уголовных дел в зависимости от номеров статей Уголовного кодекса РФ и пола судьи (рисунок 18 – 25). Данные этого среза были отфильтрованы по значению количества решений (статьи со значением количества решений меньше 30, а также типы решения со значением меньше 5 отбрасываются). Вертикальная ось диаграмм – процент от количества решений по уголовным делам, рассмотренные судьями женского и мужского пола, которые ссылаются на УК РФ. Поскольку в данный срез вошли уголовные дела всех инстанций (Кассация, Апелляция, Первая, Надзор), ссылающиеся на УК РФ, среди типов решений присутствуют Иск удовлетворен, Иск не удовлетворен и т.п.

Из среза видно, что по статье 158 (Кража) судьи женского пола вынесли больший процент решений, чем судьи мужского пола, по

следующим типам решений: Решение отменено, Прекращено за примирением сторон, Возврат дела. А судьи мужского пола вынесли больший процент по типам решений: Иск удовлетворен, Оставлено без рассмотрения.

По статье 228 (Незаконные приобретение, хранение, перевозка, изготовление, переработка наркотических средств, психотропных веществ или их аналогов, а также незаконные приобретение, хранение, перевозка растений, содержащих наркотические средства или психотропные вещества, либо их частей, содержащих наркотические средства или психотропные вещества) судьи мужского пола вынесли больший процент решений, чем судьи женского пола, по следующим типам решений: Иск удовлетворен, Производство прекращено.

По статье 161 (Грабеж) судьи женского пола вынесли больший процент решений по следующим типам решений: Возврат дела. А судьи мужского пола вынесли больший процент по типам решений: Иск удовлетворен.

По статье 111 (Умышленное причинение тяжкого вреда здоровью) процент решений между судьями разных полов не сильно отличается.

По статье 166 (Неправомерное завладение автомобилем или иным транспортным средством без цели хищения) судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено, Иск удовлетворен.

По статье 159 (Мошенничество) судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен, Решение отменено. А судьи мужского пола вынесли больший процент по типам решений: Производство прекращено, Иск удовлетворен.

По статье 264 (Нарушение правил дорожного движения и эксплуатации транспортных средств) судьи мужского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен, Оставлено без рассмотрения.

По статье 162 (Разбой) судьи мужского пола вынесли больший процент решений по следующим типам решений: Иск удовлетворен.

По статье 105 (Убийство) судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Иск удовлетворен.

По статье 308 (Отказ свидетеля или потерпевшего от дачи показаний) процент решений между судьями разных полов не сильно отличается.

По статье 222 (Незаконные приобретение, передача, сбыт, хранение, перевозка или ношение оружия, его основных частей, боеприпасов) судьи мужского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По статье 232 (Организация либо содержание притонов или систематическое предоставление помещений для потребления наркотических средств, психотропных веществ или их аналогов) процент решений между судьями разных полов не отличается.

По статье 116 (Побои) судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено, Иск удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Виновен.

По статье 318 (Применение насилия в отношении представителя власти) процент решений между судьями разных полов не сильно отличается.

По статье 119 (Угроза убийством или причинением тяжкого вреда здоровью) судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По статье 112 (Умышленное причинение средней тяжести вреда здоровью) судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Производство прекращено.

По статье 260 (Незаконная рубка лесных насаждений) судьи мужского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен.

По статье 160 (Присвоение или растрата) судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По статье 306 (Заведомо ложный донос) судьи мужского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По статье 115 (Умышленное причинение легкого вреда здоровью) судьи мужского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По статье 167 (Умышленные уничтожение или повреждение имущества) судьи мужского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По статье 307 (Заведомо ложные показания, заключение эксперта, специалиста или неправильный перевод) процент решений между судьями разных полов не сильно отличается.

По статье 223 (Незаконное изготовление оружия) судьи мужского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По статье 234 (Незаконный оборот сильнодействующих или ядовитых веществ в целях сбыта) процент решений между судьями разных полов не сильно отличается.

Статья 130 утратила силу.

По статье 139 (Нарушение неприкосновенности жилища) судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Виновен.

По статье 327 (Подделка, изготовление или сбыт поддельных документов, государственных наград, штампов, печатей, бланков) судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено. А судьи мужского пола вынесли больший процент по типам решений: Виновен.

По статье 337 (Самовольное оставление части или места службы) процент решений между судьями разных полов не отличается.

По статье 109 (Причинение смерти по неосторожности) судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По статье 131 (Изнасилование) судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Виновен.

По статье 163 (Вымогательство) судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Виновен.

По статьям 319 (Оскорбление представителя власти), 286 (Превышение должностных полномочий), 146(Нарушение авторских и смежных прав) процент решений между судьями разных полов не отличается.

Рисунок 21 – Распределение типов решений уголовных дел в зависимости от номеров статей УК РФ и пола судьи, нормированная гистограмма с накоплением

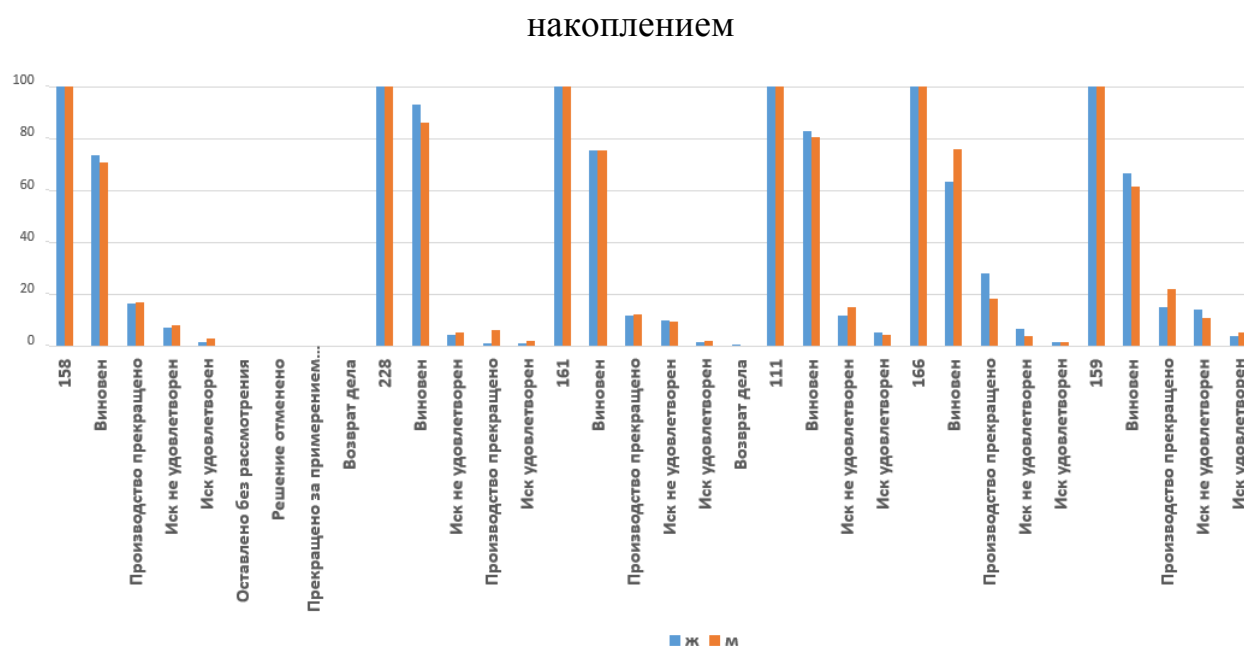


Рисунок 22 – Распределение типов решений уголовных дел в зависимости от номеров статей УК РФ и пола судьи, гистограмма с группировкой

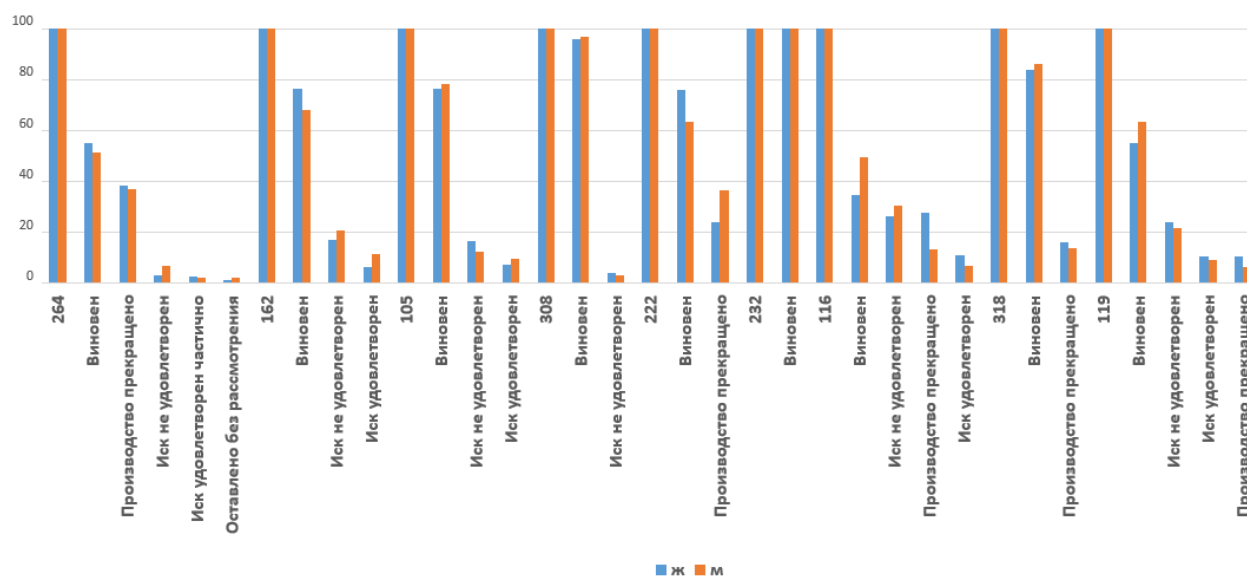


Рисунок 23 – Распределение типов решений уголовных дел в зависимости от номеров статей УК РФ и пола судьи, гистограмма с группировкой

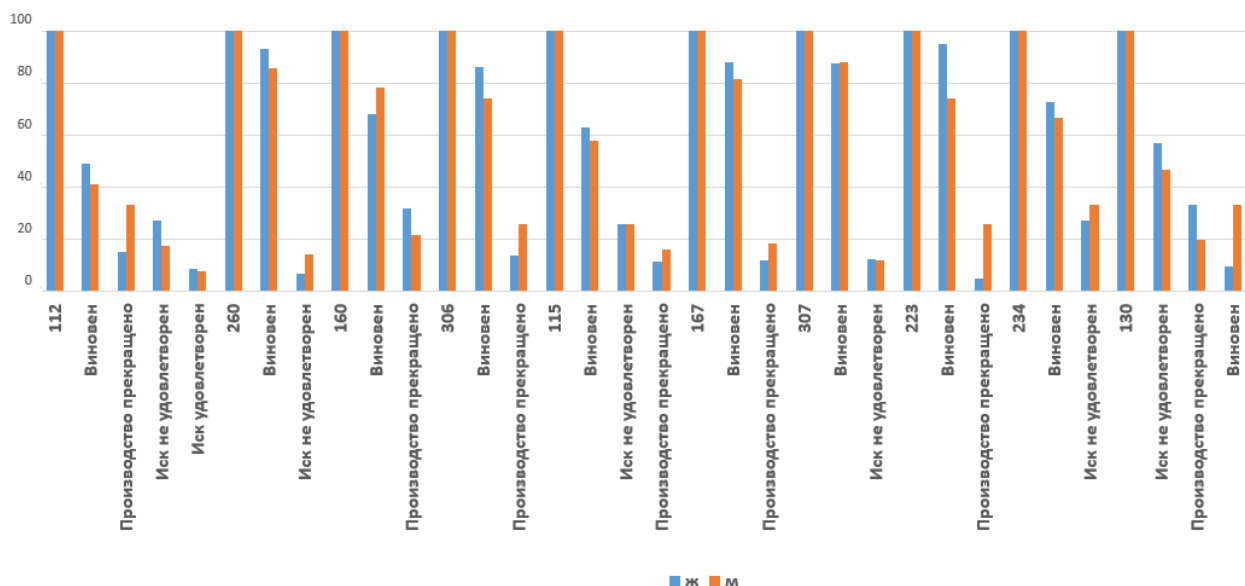


Рисунок 24 – Распределение типов решений уголовных дел в зависимости от номеров статей УК РФ и пола судьи, гистограмма с группировкой

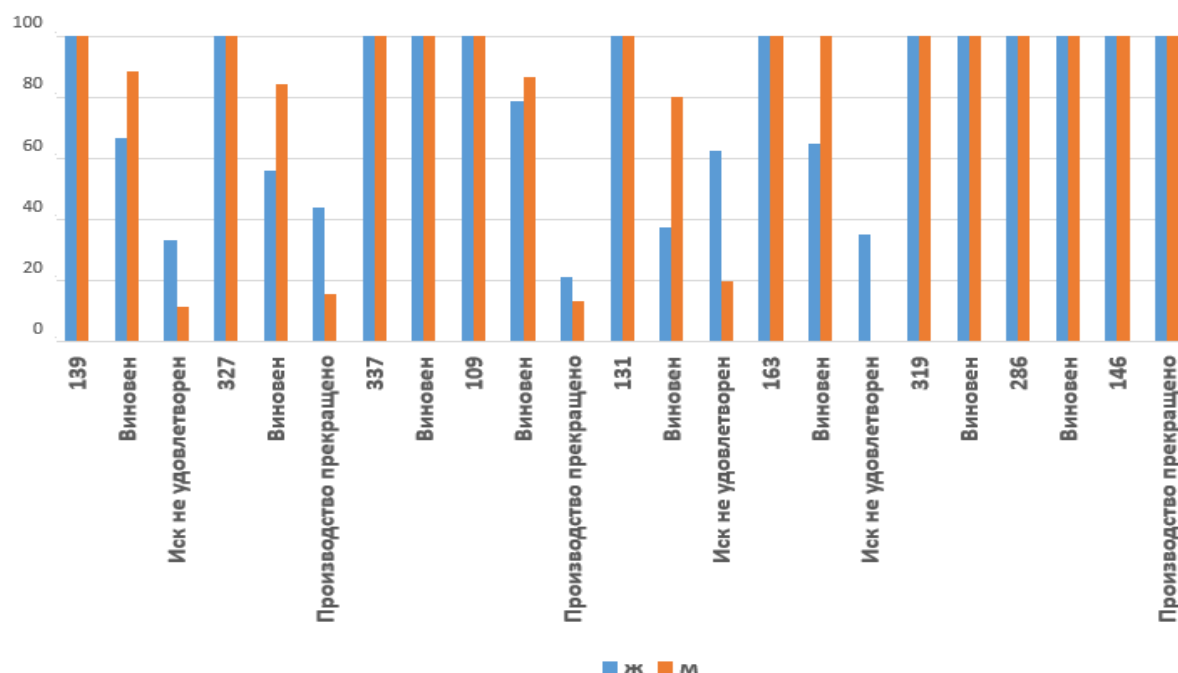


Рисунок 25 – Распределение типов решений уголовных дел в зависимости от номеров статей УК РФ и пола судьи, гистограмма с группировкой

11. Распределение решений уголовных дел в зависимости от глав статей Уголовного кодекса РФ (рисунок 26). Вертикальная ось диаграммы – процент от количества решений по уголовным делам, которые ссылаются на УК РФ. Из среза видно, что большая часть решений по уголовным делам, ссылающихся на УК РФ, приходится на главу «Преступления против собственности» (около 60%) [4].

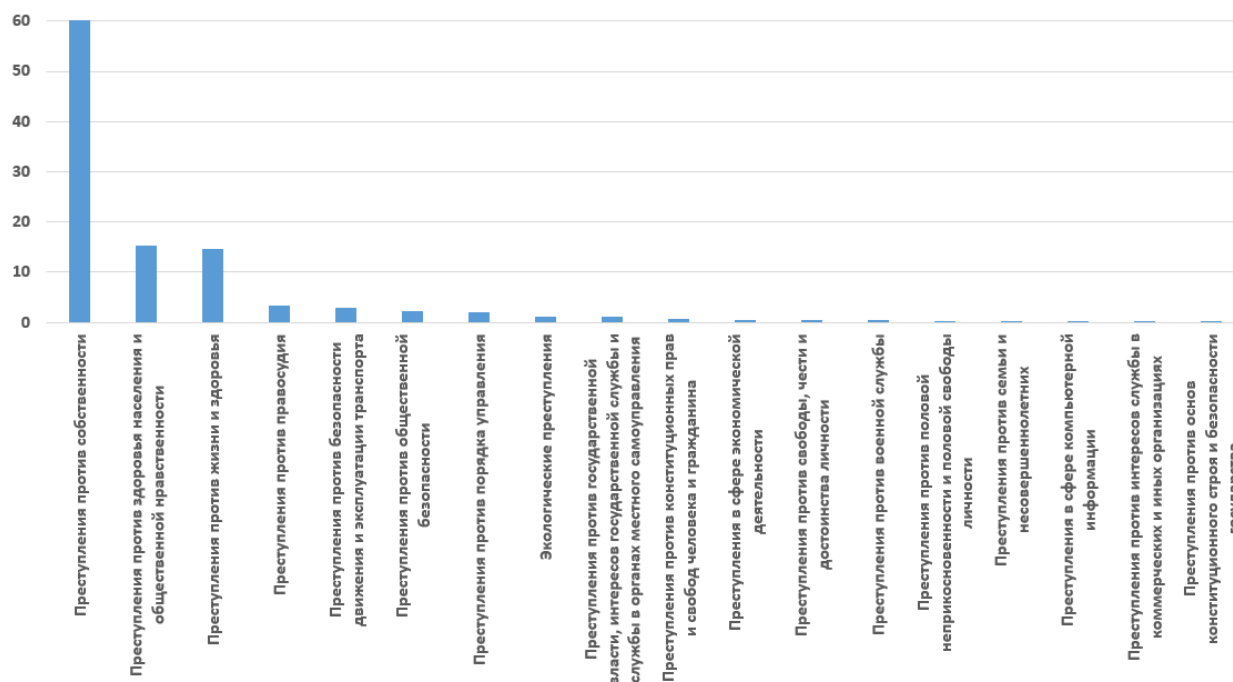


Рисунок 26 – Распределение решений уголовных дел в зависимости от глав статей УК РФ

12. Распределение типов решений уголовных дел в зависимости от глав статей Уголовного кодекса РФ и пола судьи (рисунок 27 – 32). Данные этого среза были отфильтрованы по значению количества решений (главы со значением количества решений меньше 10, а также типы решения со значением меньше 5 отбрасываются). Вертикальная ось диаграмм – процент от количества решений по уголовным делам, рассмотренные судьями женского и мужского пола, которые ссылаются на УК РФ. Поскольку в данный срез вошли уголовные дела всех инстанций (Кассация, Апелляция, Первая, Надзор), ссылающиеся на УК РФ, среди типов решений присутствуют Иск удовлетворен, Иск не удовлетворен и т.п.

Из среза видно, что по главе «Преступления против собственности» судьи женского пола вынесли больший процент решений, чем судьи мужского пола, по следующим типам решений: Возврат дела, Решение отменено, Прекращено за примирением сторон. А судьи мужского пола вынесли больший процент по типам решений: Иск удовлетворен, Оставлено без рассмотрения, Передать дело.

По главе «Преступления против здоровья населения и общественной нравственности» судьи мужского пола вынесли больший процент решений по следующим типам решений: Производство прекращено, Иск удовлетворен.

По главе «Преступления против жизни и здоровья» судьи женского пола вынесли больший процент решений по следующим типам решений: Передать дело, Возврат дела.

По главе «Преступления против правосудия» судьи мужского пола вынесли больший процент решений по следующим типам решений: Производство прекращено, Иск удовлетворен.

По главе «Преступления против безопасности движения и эксплуатации транспорта» судьи мужского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен, Оставлено без рассмотрения.

По главе «Преступления против общественной безопасности» судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Производство прекращено.

По главе «Преступления против порядка управления» судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено. А судьи мужского пола вынесли больший процент по типам решений: Иск удовлетворен.

По главе «Экологические преступления» судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По главе «Преступления против государственной власти, интересов государственной службы и службы в органах местного самоуправления» судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По главе «Преступления против конституционных прав и свобод человека и гражданина» судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Виновен.

По главе «Преступления в сфере экономической деятельности» судьи женского пола вынесли больший процент решений по следующим типам решений: Виновен. А судьи мужского пола вынесли больший процент по типам решений: Иск не удовлетворен.

По главе «Преступления против свободы, чести и достоинства личности» судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен, Производство прекращено. А судьи мужского пола вынесли больший процент по типам решений: Виновен, Иск удовлетворен.

По главе «Преступления против военной службы» процент решений между судьями разных полов не отличается.

По главе «Преступления против половой неприкосновенности и половой свободы личности» судьи женского пола вынесли больший процент решений по следующим типам решений: Иск не удовлетворен. А судьи мужского пола вынесли больший процент по типам решений: Виновен.

По главе «Преступления против семьи и несовершеннолетних» процент решений между судьями разных полов не отличается.

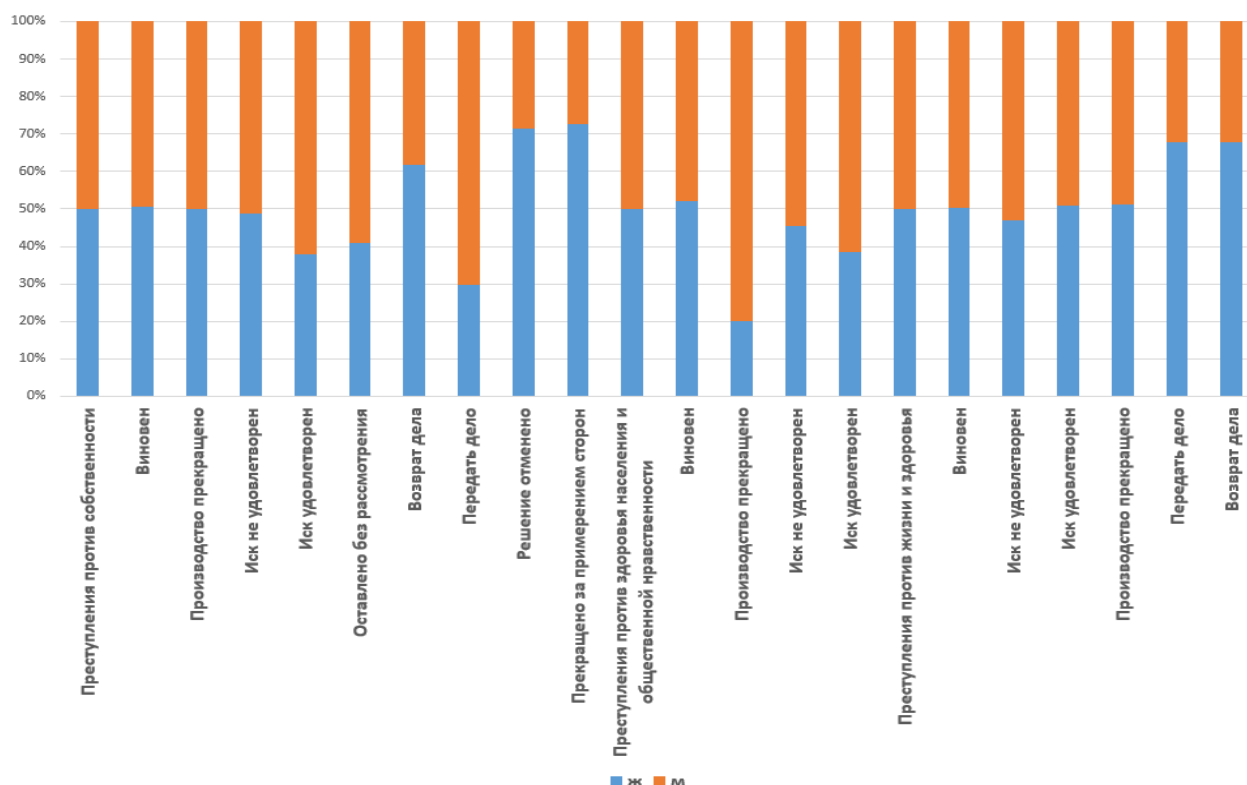


Рисунок 27 – Распределение типов решений уголовных дел в зависимости от глав статей УК РФ и пола судьи, нормированная гистограмма с накоплением

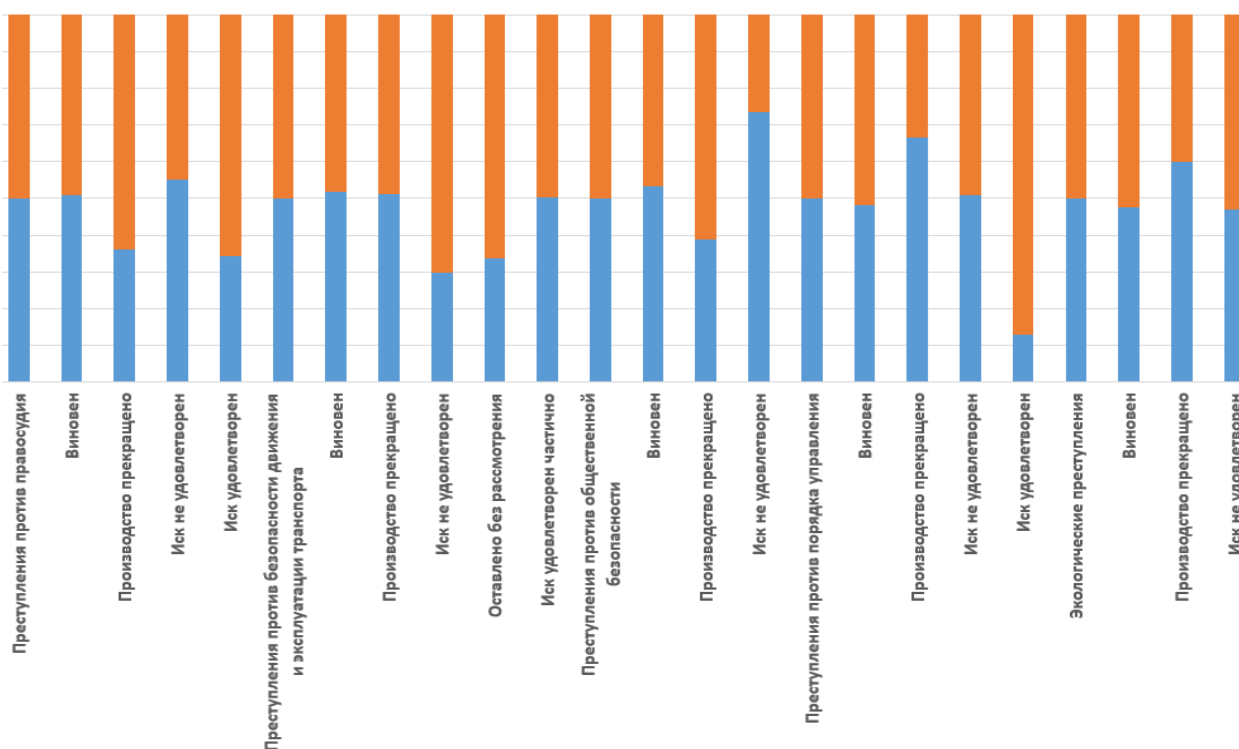


Рисунок 28 – Распределение типов решений уголовных дел в зависимости от глав статей УК РФ и пола судьи, нормированная гистограмма с накоплением

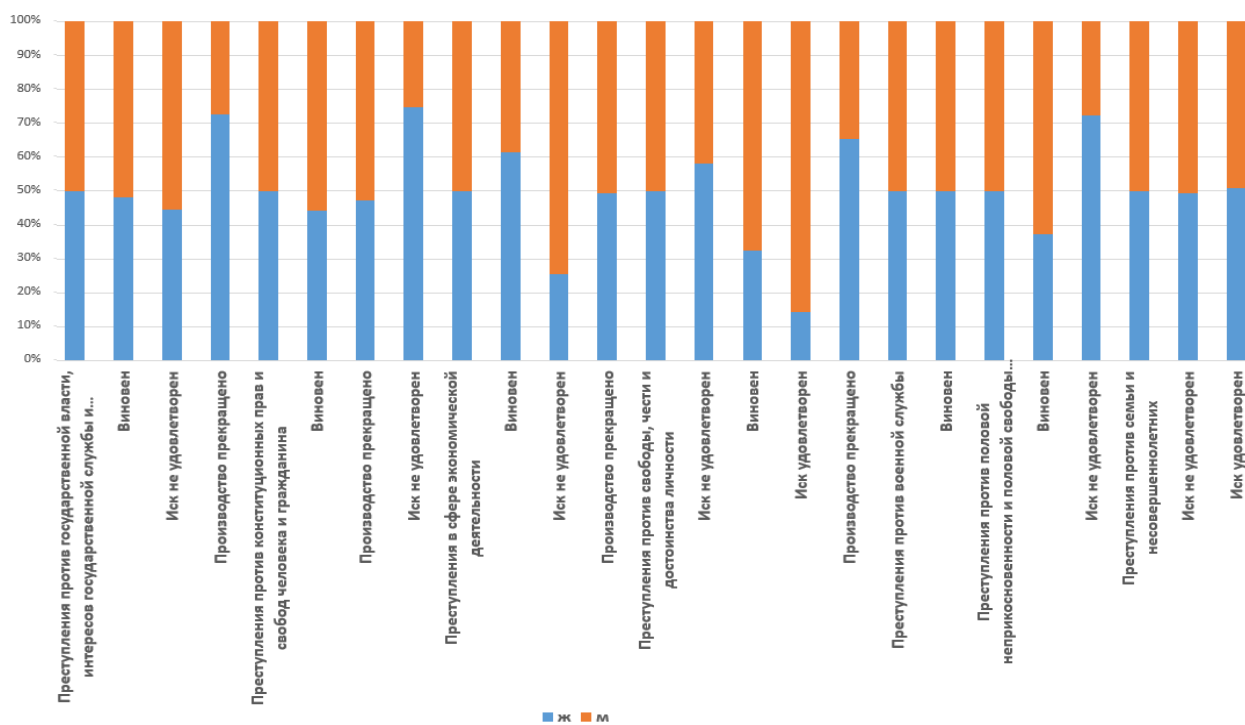


Рисунок 29 – Распределение типов решений уголовных дел в зависимости от глав статей УК РФ и пола судьи, нормированная гистограмма с накоплением

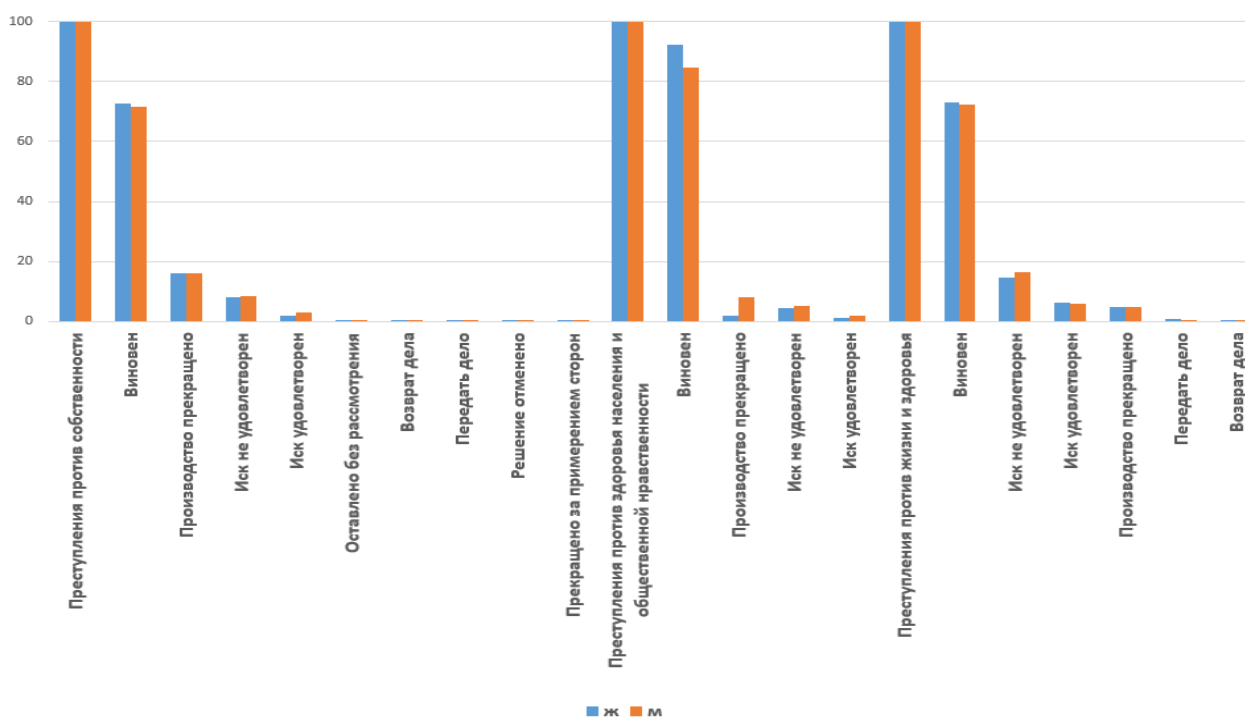


Рисунок 30 – Распределение типов решений уголовных дел в зависимости от глав статей УК РФ и пола судьи, гистограмма с группировкой

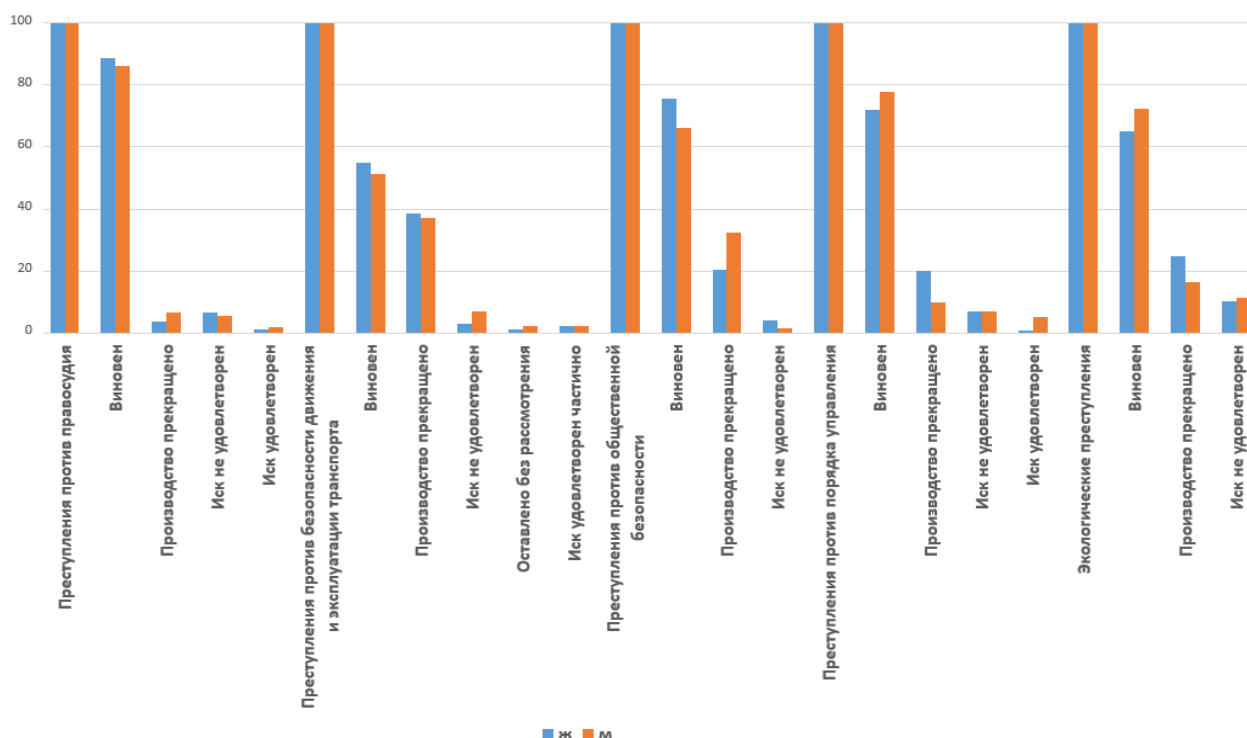


Рисунок 31 – Распределение типов решений уголовных дел в зависимости от глав статей УК РФ и пола судьи, гистограмма с группировкой

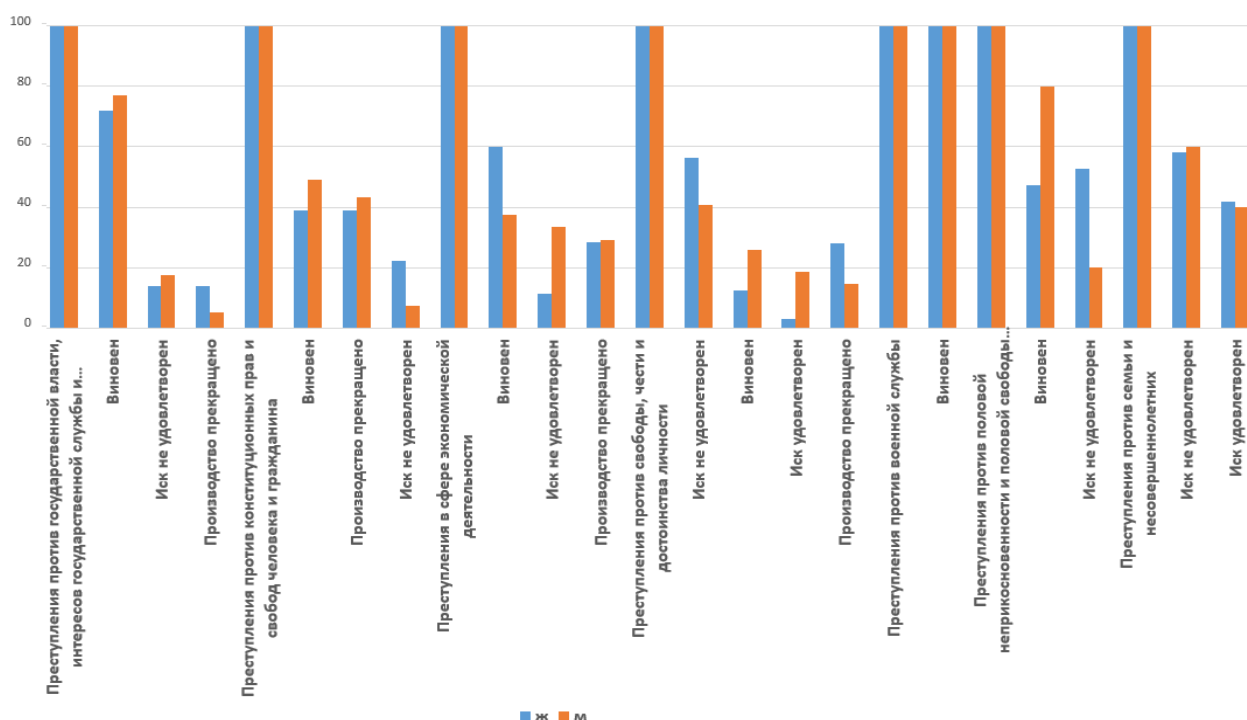


Рисунок 32 – Распределение типов решений уголовных дел в зависимости от глав статей УК РФ и пола судьи, гистограмма с группировкой

13. Распределение типов решений уголовных дел первой инстанции в зависимости от глав статей УК РФ и пола судьи (рисунок 33 – 36). Данные этого среза были отфильтрованы по значению количества решений (главы со

значением количества решений меньше 10). Вертикальная ось диаграмм – процент от количества решений по уголовным делам первой инстанции, рассмотренные судьями женского и мужского пола, которые ссылаются на УК РФ.

Из среза видно, что по главе «Преступления против собственности» судьи женского пола вынесли больший процент решений, чем судьи мужского пола, по следующим типам решений: Возврат дела, Прекращено за примирением сторон. А судьи мужского пола вынесли больший процент по типам решений: Оставлено без рассмотрения.

По главе «Преступления против здоровья населения и общественной нравственности» судьи женского пола вынесли больший процент решений по следующим типам решений: Передать дело. А судьи мужского пола вынесли больший процент по типам решений: Производство прекращено.

По главе «Преступления против жизни и здоровья» судьи женского пола вынесли больший процент решений по следующим типам решений: Передать дело, Возврат дела.

По главе «Преступления против правосудия» судьи мужского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По главе «Преступления против безопасности движения и эксплуатации транспорта» судьи мужского пола вынесли больший процент решений по следующим типам решений: Оставлено без рассмотрения.

По главе «Преступления против общественной безопасности» судьи женского пола вынесли больший процент решений по следующим типам решений: Виновен. А судьи мужского пола вынесли больший процент по типам решений: Производство прекращено.

По главе «Преступления против порядка управления» судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено. А судьи мужского пола вынесли больший процент по типам решений: Виновен.

По главе «Экологические преступления» судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено. А судьи мужского пола вынесли больший процент по типам решений: Возврат дела.

По главе «Преступления против государственной власти, интересов государственной службы и службы в органах местного самоуправления» судьи женского пола вынесли больший процент решений по следующим типам решений: Производство прекращено.

По главе «Преступления против конституционных прав и свобод человека и гражданина» судьи женского пола вынесли больший процент решений по следующим типам решений: Виновен.

По главам «Преступления против военной службы», «Преступления в сфере экономической деятельности», «Преступления против половой неприкосновенности и половой свободы личности» процент решений между судьями разных полов не отличается.

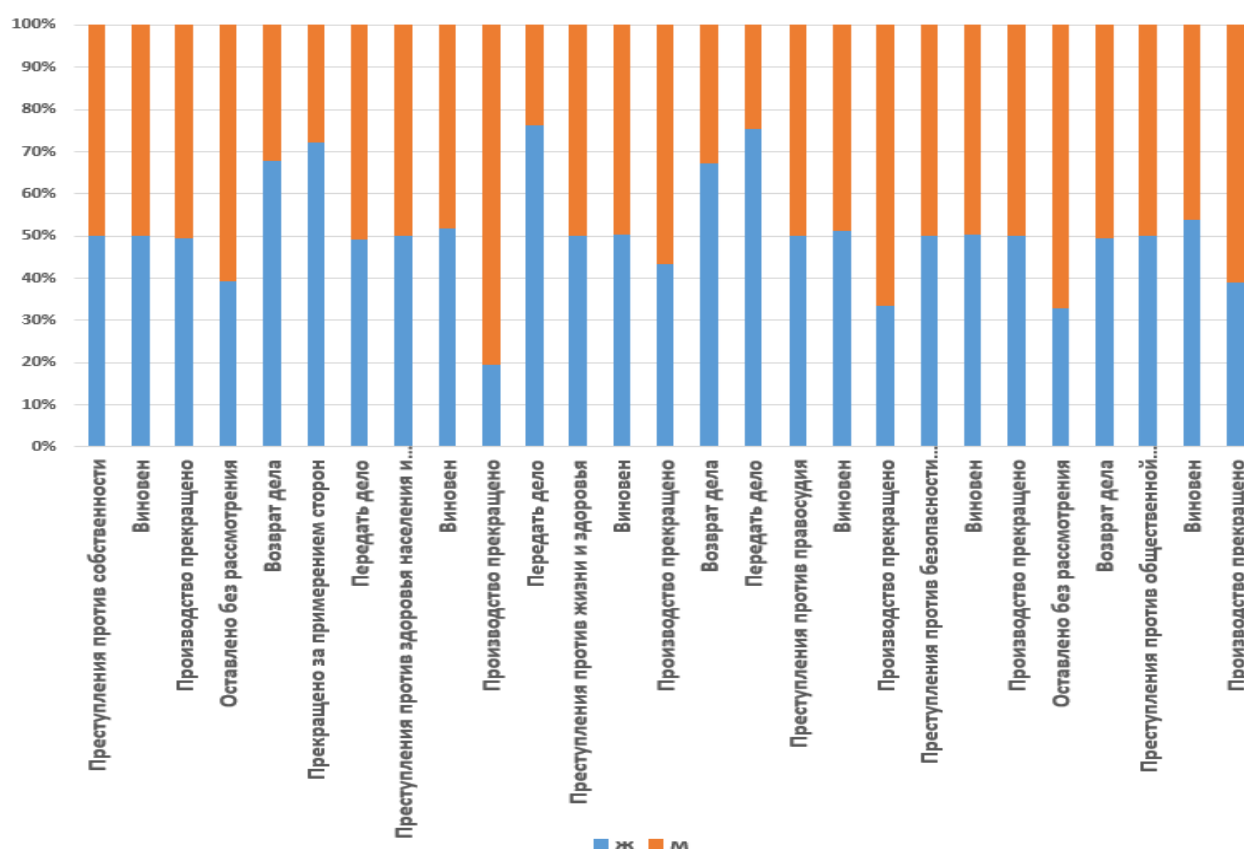


Рисунок 33 – Распределение типов решений уголовных дел первой инстанции в зависимости от глав статей УК РФ и пола судьи, нормированная гистограмма с накоплением

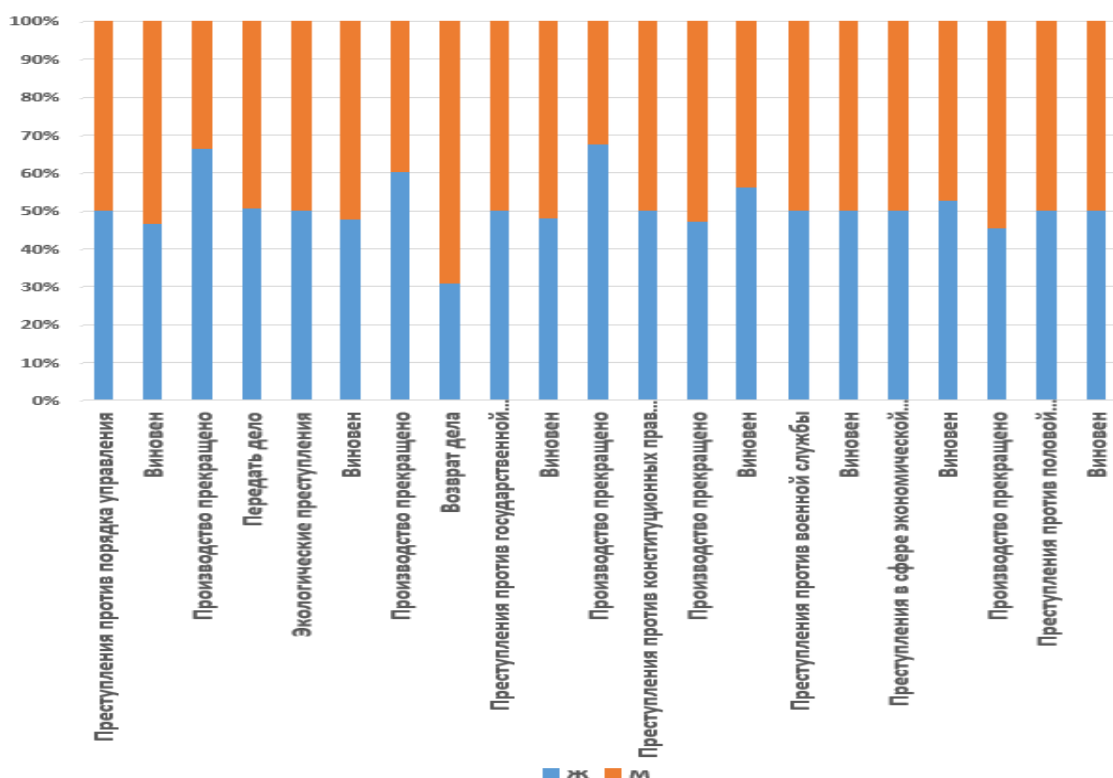


Рисунок 34 – Распределение типов решений уголовных дел первой инстанции в зависимости от глав статей УК РФ и пола судьи, нормированная гистограмма с накоплением

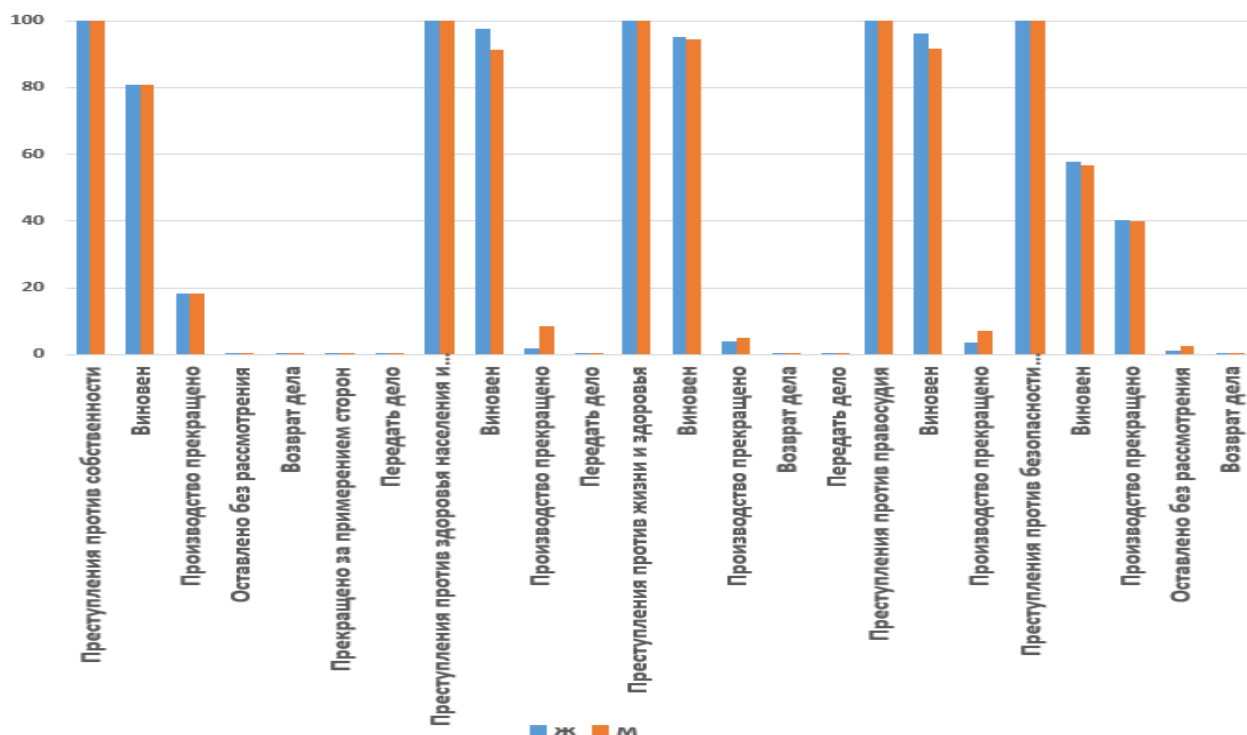


Рисунок 35 – Распределение типов решений уголовных дел первой инстанции в зависимости от глав статей УК РФ и пола судьи, нормированная гистограмма с накоплением

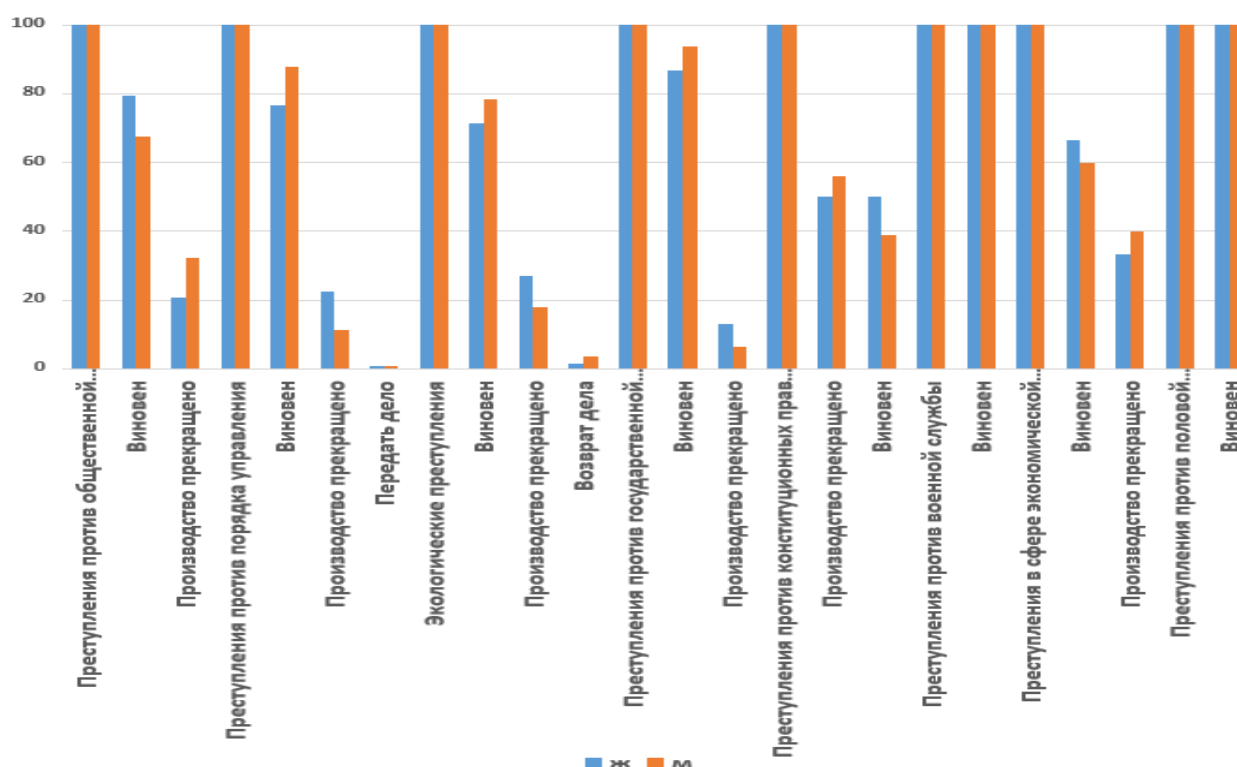


Рисунок 36 – Распределение типов решений уголовных дел первой инстанции в зависимости от глав статей УК РФ и пола судьи, нормированная гистограмма с накоплением

14. Распределение типов решений по судьям (рисунок 37 – 40). Данные этого среза были отфильтрованы по значению количества решений (судьи со значением количества решений меньше 100 отбрасываются). Вертикальная ось диаграмм – процент от количества решений, рассмотренные соответствующими судьями.

Тип решения Иск не удовлетворен наиболее высок у судьи Кривошеин Юрий Гаврилович и составляет почти 90% от всего объема рассмотренных им решений.

Тип решения Иск удовлетворен наиболее высок у судьи Мацкевич Рита Николаевна и составляет больше 60%.

Тип решения Иск удовлетворен частично наиболее высок у судьи Юровский Игорь Петрович и составляет 23%.

Тип решения Виновен наиболее высок у судьи Матыскина Л. С. и составляет больше 80%.

Тип решения Оставлено без рассмотрения наиболее высок у судьи Типигин Сергей Иванович и составляет чуть больше 9%.

Тип решения Передать дело наиболее высок у судьи Тыняный Александр Михайлович и составляет почти 5%.

Тип решения Производство прекращено наиболее высок у судьи Васильев Юрий Михайлович 26%.

Кроме того, можно рассмотреть конкретного судью, к каким типам решений он тяготеет. Так, у Родникова А. А.: 41% - Иск не удовлетворен, 5% - Иск удовлетворен частично, 15% - Виновен, 28% - Иск удовлетворен и 10% - Производство прекращено. Вотина Виктория Ивановна: 51% - Иск не удовлетворен, 5% - Иск удовлетворен частично, 4% - Виновен, 30% - Иск удовлетворен и 5% - Производство прекращено. Глумова Татьяна Анатольевна: 28% - Иск не удовлетворен, 4% - Иск удовлетворен частично, 25% - Виновен, 27% - Иск удовлетворен и 11% - Производство прекращено.

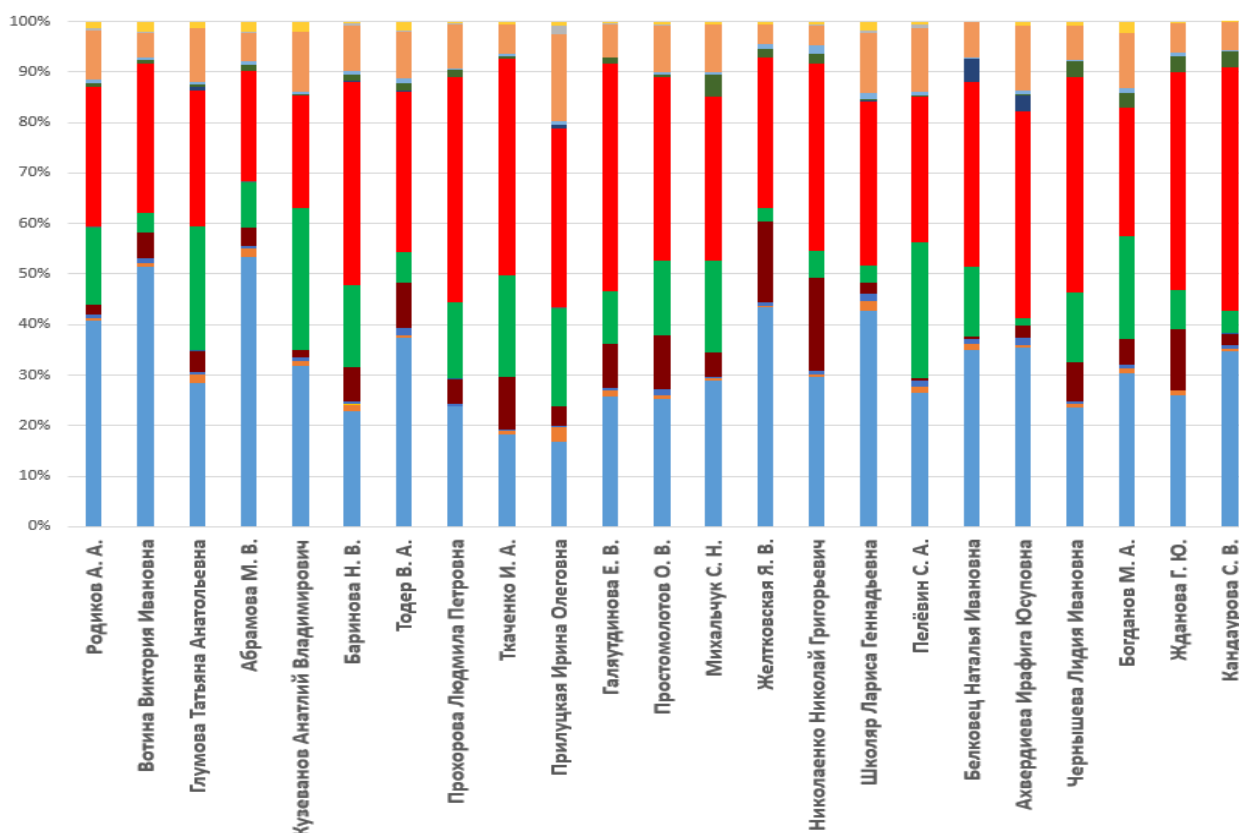


Рисунок 37 – Распределение типов решений по судьям

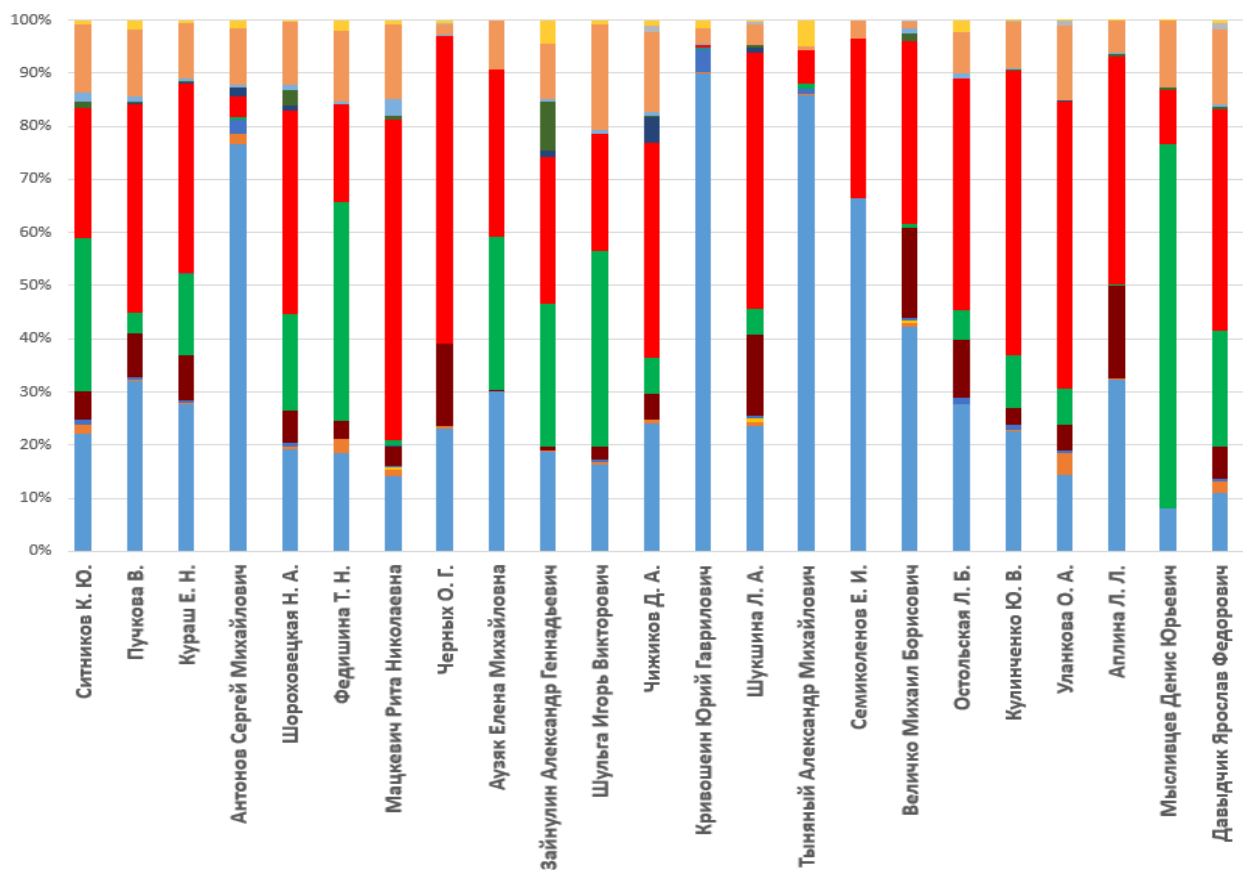


Рисунок 38 – Распределение типов решений по судьям

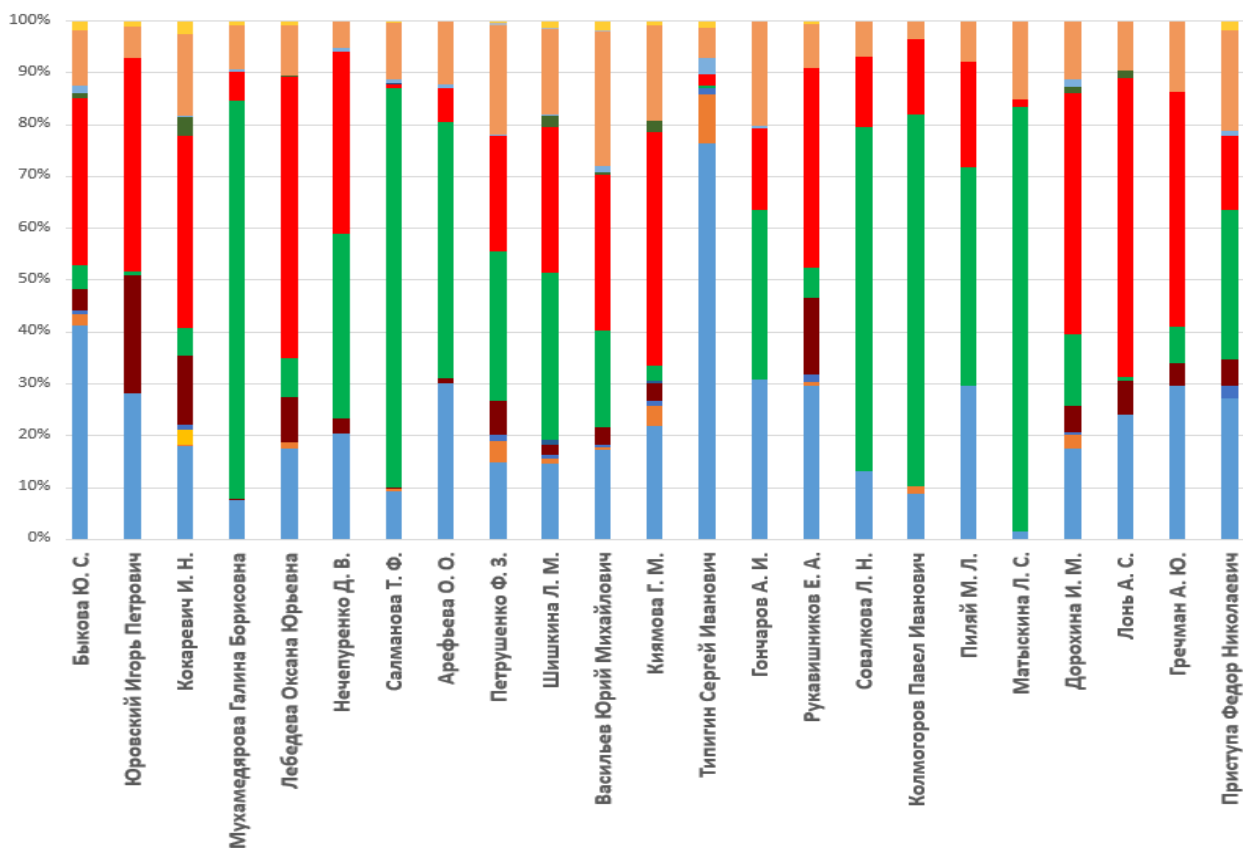


Рисунок 39 – Распределение типов решений по судьям



Рисунок 40 – Легенда распределения типов решений по судьям

15. Распределение вида дела по судьям (рисунок 41 – 43). Данные этого среза были отфильтрованы по значению количества решений (судьи со значением количества решений меньше 100 отбрасываются). Вертикальная ось диаграмм – процент от количества решений, рассмотренные соответствующими судьями.

Как можно заметить, большая часть судей рассматривает дела различного вида, хоть и в различных пропорциях. Например, Родников А. А. 35% - Уголовное, 42% - Гражданское, 14% - По материалам, 9% - Административное. Однако есть и такие судьи, которые рассматривают дела лишь одного типа. Например, Кривошеин Юрий Гаврилович и Тыняный Александр Михайлович рассматривают лишь уголовные дела. Либо, у некоторых судей преобладают дела одного вида, например, у Желтковской Я. В. это гражданские дела.

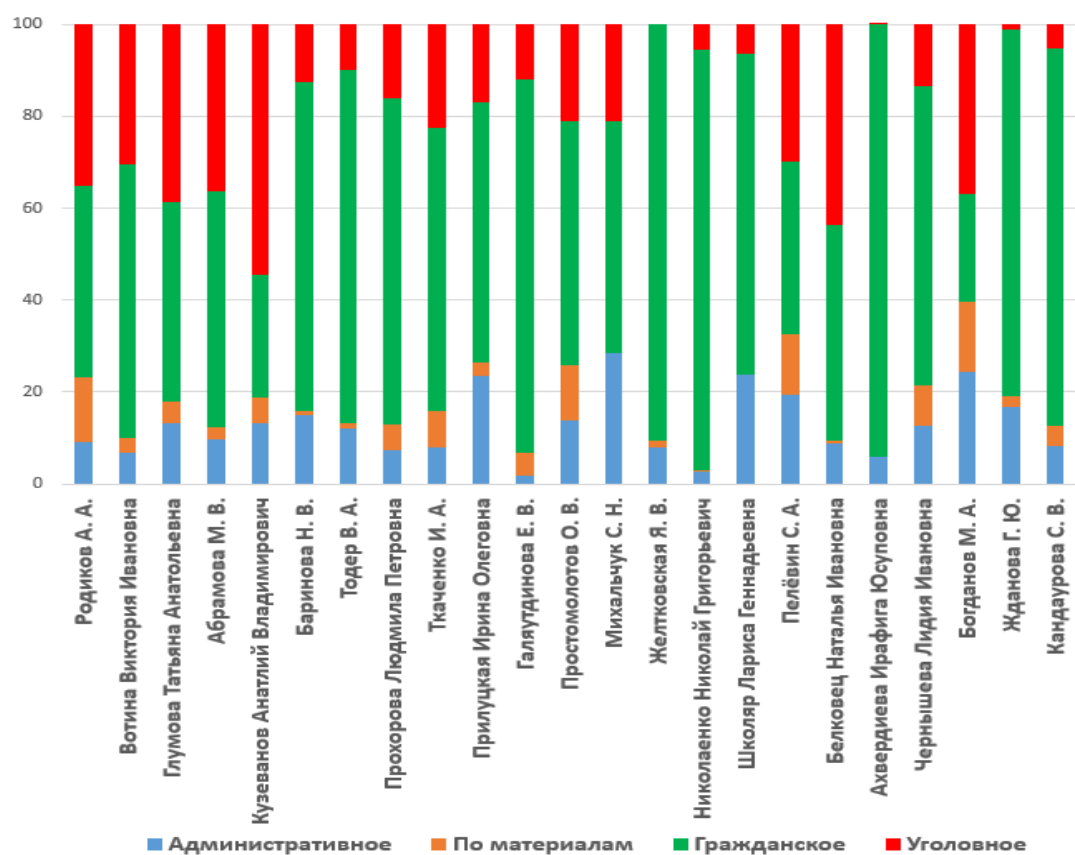


Рисунок 41 – Распределение видов дел по судьям

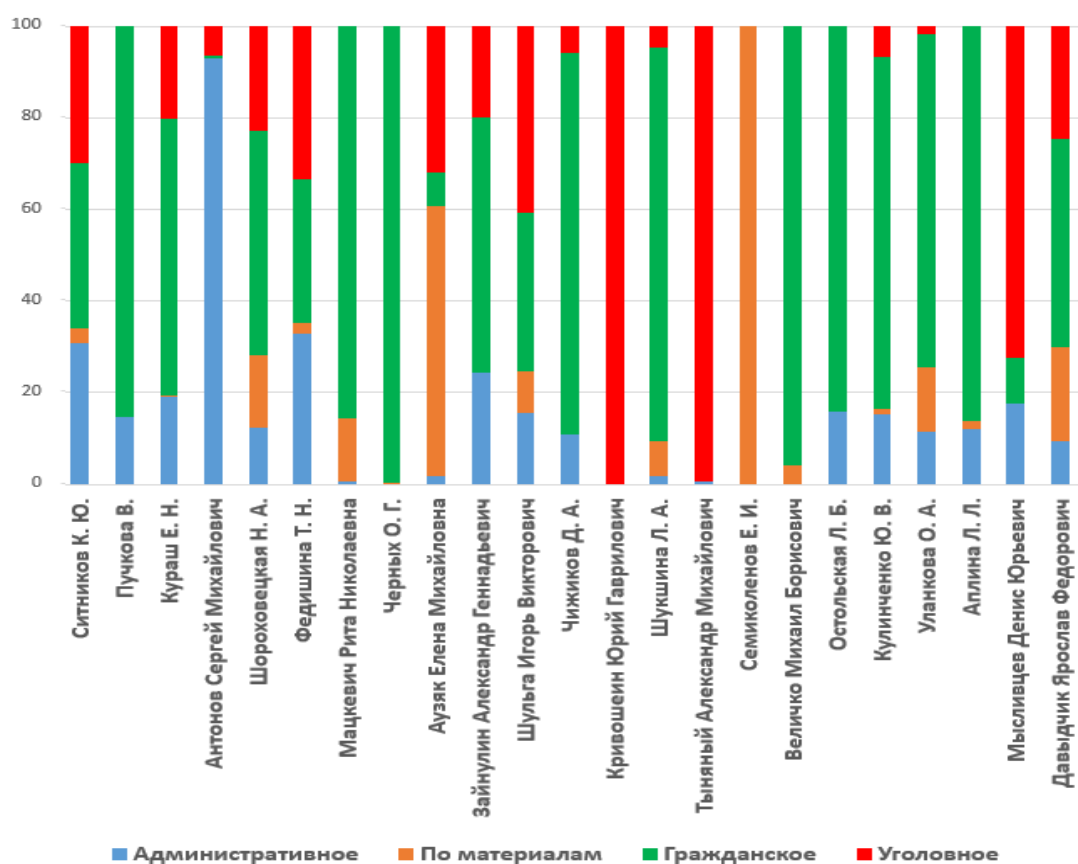


Рисунок 42 – Распределение видов дел по судьям

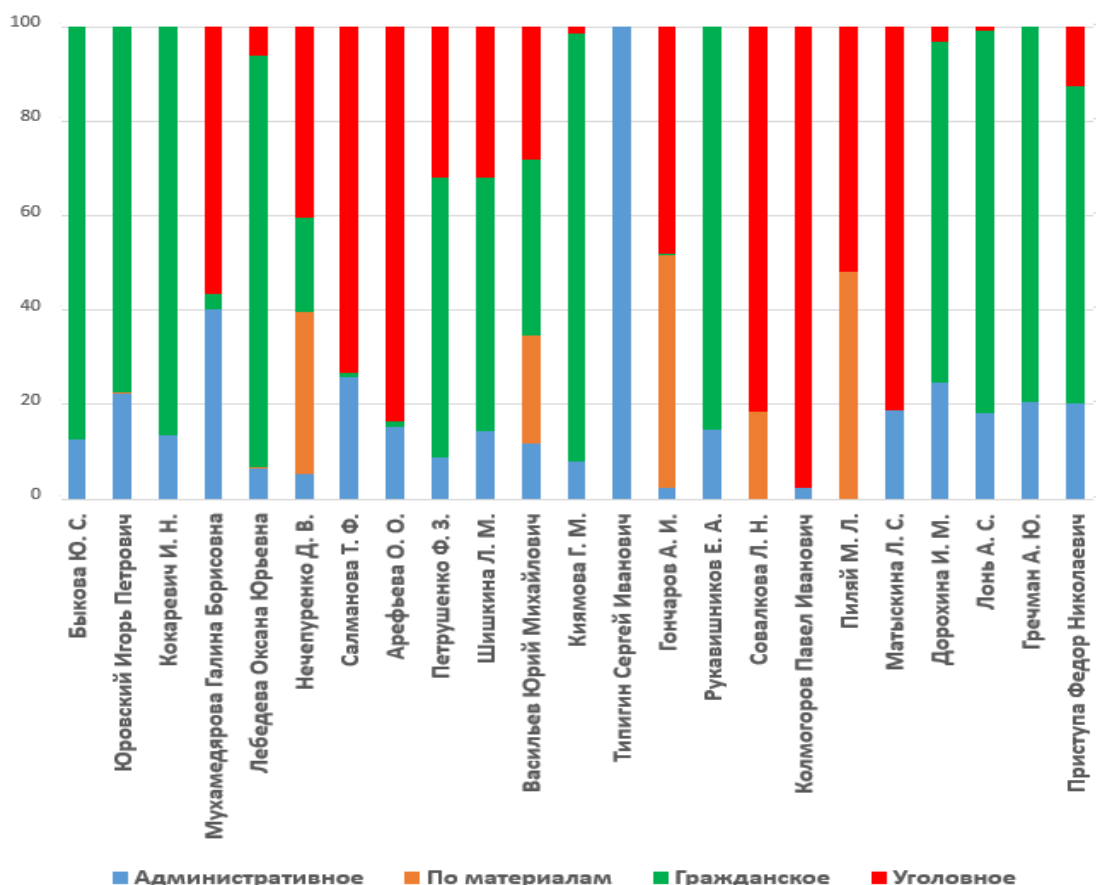


Рисунок 43 – Распределение видов дел по судьям

16. Распределение типов решения по адвокатам (рисунок 43 – 45). Данные этого среза были отфильтрованы по значению количества решений (адвокаты со значением количества решений меньше 100 отбрасываются). Вертикальная ось диаграмм – процент от количества решений, соответствующих адвокатам.

Тип решения Иск не удовлетворен наиболее высок у адвоката Гордиенко О.Ю. и составляет чуть больше 87% от всего объема решений адвоката.

Тип решения Иск удовлетворен наиболее высок у адвоката Герман Е.С. и составляет чуть больше 67%.

Тип решения Иск удовлетворен частично наиболее высок у адвоката Ярошинская Г.Г. и составляет 14%.

Тип решения Виновен наиболее высок у адвоката Чарного С.М. и составляет почти 88%.

Тип решения Передать дело наиболее высок у адвоката Учуткина И.Ю. и составляет почти 10%.

Тип решения Производство прекращено наиболее высок у адвоката Карабец Ю.С. и составляет чуть больше 16%.

Кроме того, можно рассмотреть конкретного адвоката, какие типы решений преобладают. Так, у Ульяновой А. А.: 64% - Иск не удовлетворен, 3% - Иск удовлетворен частично, 11% - Виновен, 14% - Иск удовлетворен и 4% - Производство прекращено. Факова А.В.: 35% - Иск не удовлетворен, 5% - Иск удовлетворен частично, 32% - Виновен, 16% - Иск удовлетворен и 9% - Производство прекращено. Булыгина Г.В.: 58% - Иск не удовлетворен, 0,3% - Иск удовлетворен частично, 7% - Виновен, 28% - Иск удовлетворен и 4% - Производство прекращено.

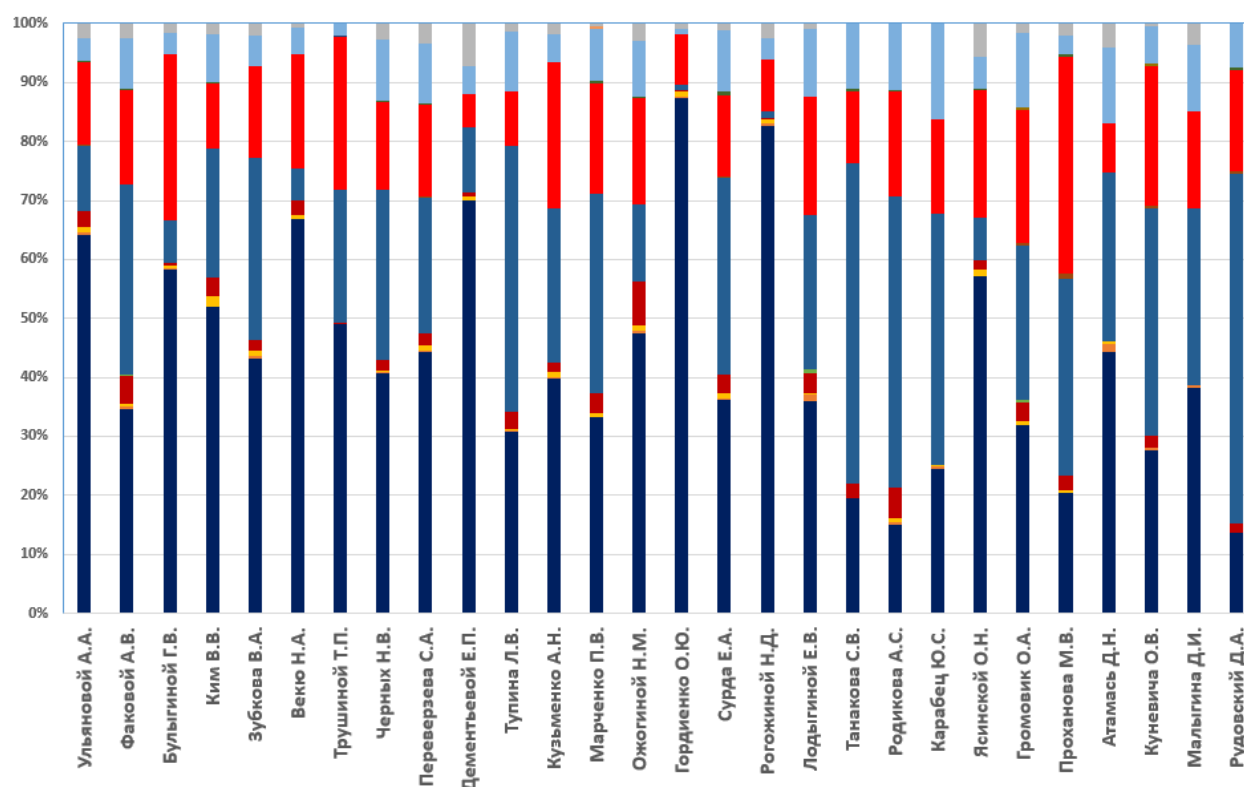


Рисунок 43 – Распределение типов решения по адвокатам

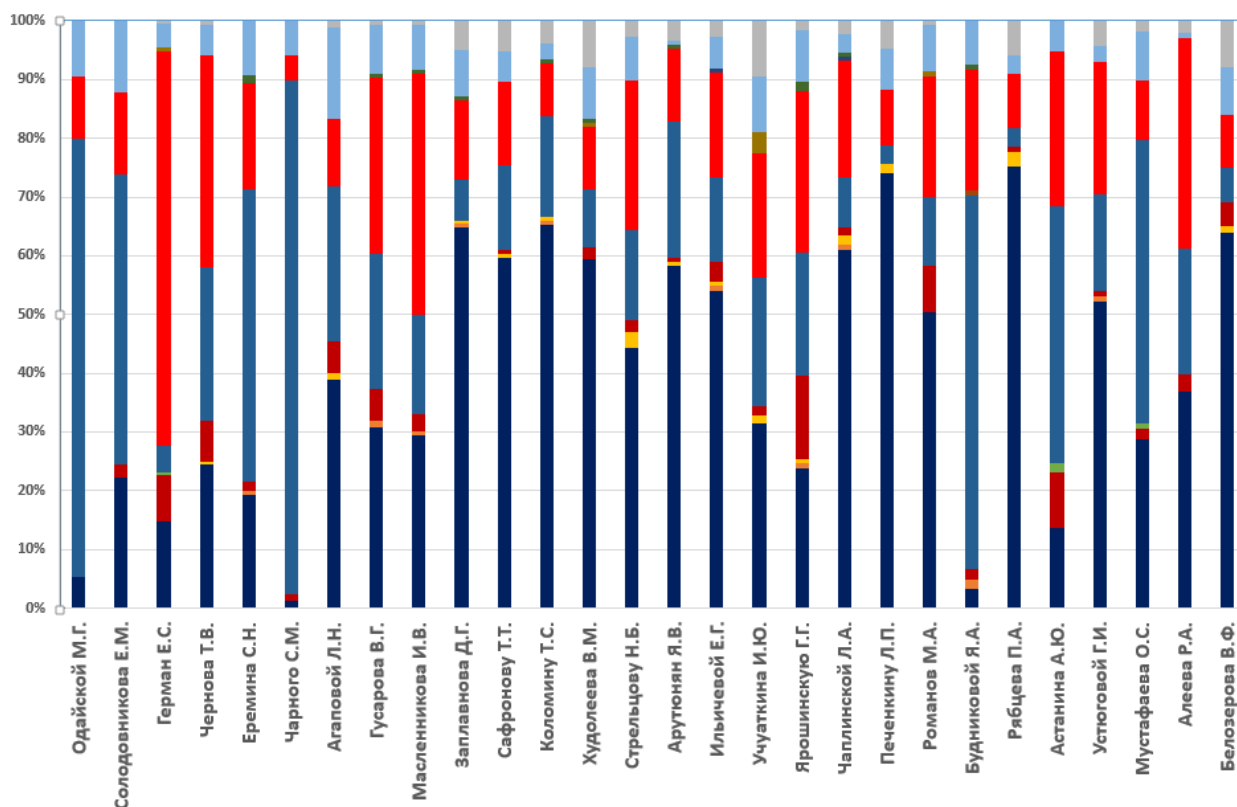


Рисунок 44 – Распределение типов решения по адвокатам



Рисунок 45 – Легенда распределения типов решения по адвокатам

17. Распределение категорий решений по адвокатам (рисунок 46 – 47). Решения по делам были разделены на три категории: условно положительные, условно отрицательные и условно нейтральные. Данные этого среза были отфильтрованы по значению количества решений (адвокаты со значением количества решений меньше 100 отбрасываются). Вертикальная ось диаграмм – процент от количества решений,

соответствующих адвокатов. Сортировка по убыванию выполнена по значениям процента положительных дел.

Так, наиболее эффективным является адвокат Герман Е.С. у которого чуть меньше 80% положительных решений по судебным делам. Также выше 50% положительных решений имеют Масленникова И.В. и Ярошинская Г.Г.

А менее эффективными являются Дементьева Е.П., Чарной С.М., Гордиенко О.Ю. у которых чуть больше 10% положительных решений, а отрицательных свыше 80%.

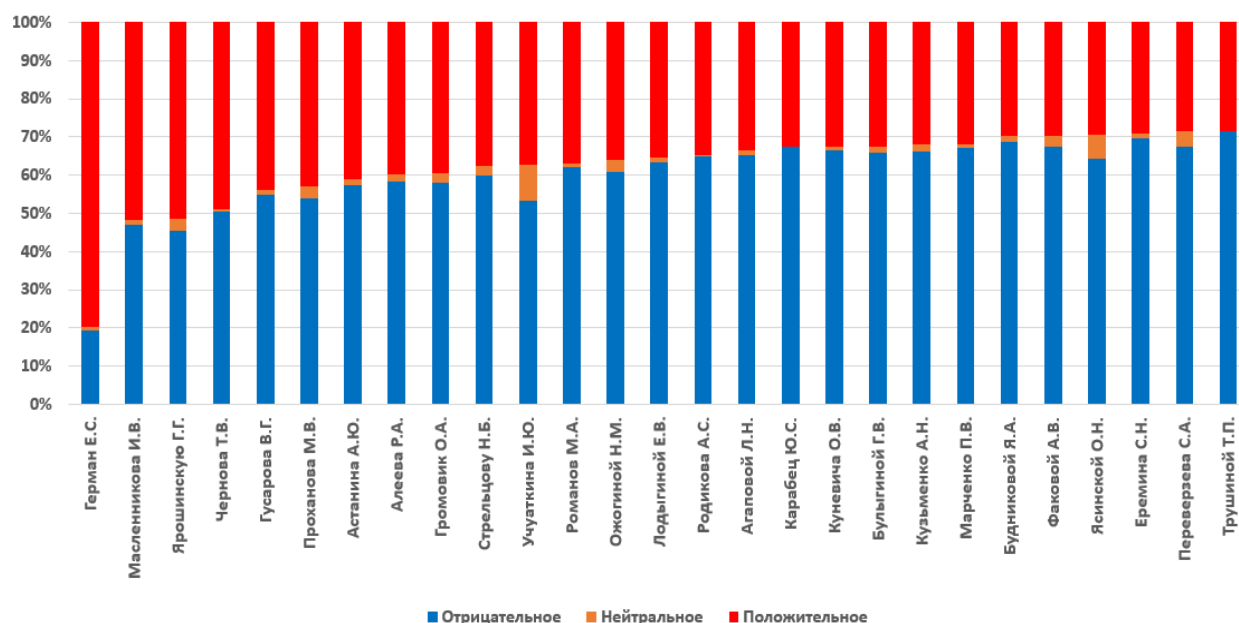


Рисунок 46 –Распределение категорий решений по адвокатам

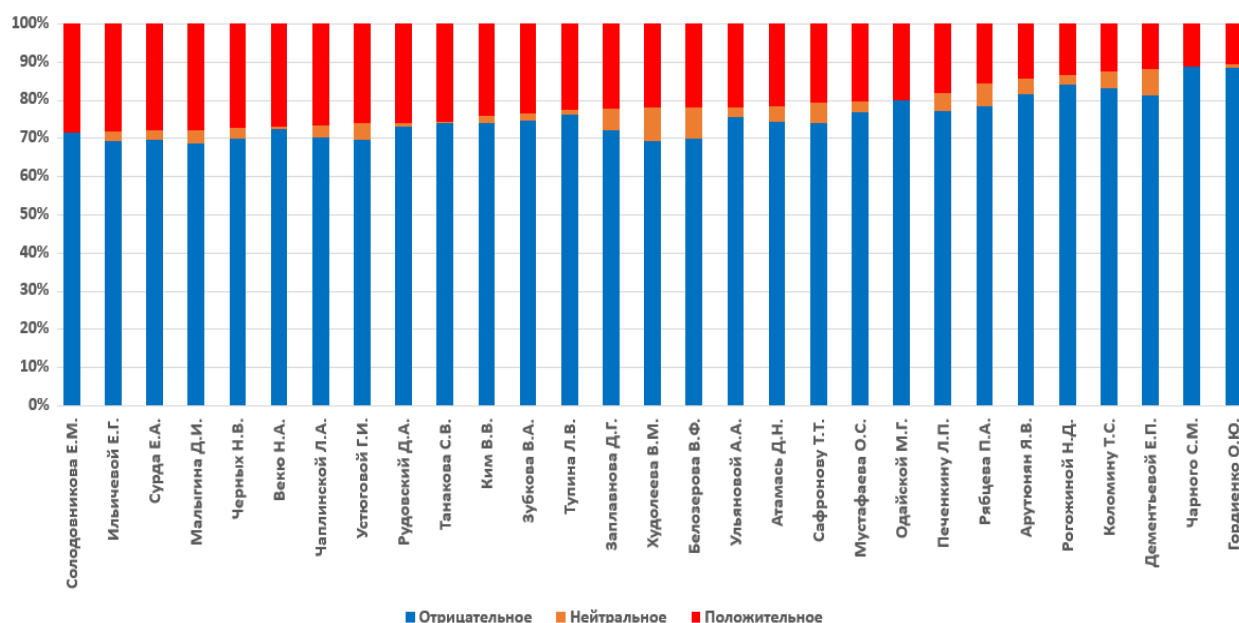


Рисунок 47 –Распределение категорий решений по адвокатам

На основе приведенных срезов можно сделать вывод, что судьи женского пола более склонны наказывать взысканием, в то время как судьи мужского пола склонны к аресту. Кроме того, прослеживаются некоторые различия при вынесении решений по главам и статьям УК РФ судьями женского и мужского полов. Например, можно сказать, что по уголовным делам главы «Преступления против половой неприкосновенности и половой свободы личности» судьи женского пола вынесли больше решений по не удовлетворению исков, а судьи мужского пола больше вынесли решений о признании виновным. Или по уголовным делам первой инстанции по статьям главы «Преступления против общественной безопасности» судьи женского пола, в отличие от судей мужского пола, выносят больше решений о признании вины. Таким образом мы можем предположить, что все же некая связь между полом судьи и типом решения и наказания существует.

Более того, на представленных срезах, можно увидеть какие типы решений и виды дел преобладают у того или иного судьи. Так же, можно посмотреть на эффективность работы адвокатов, то есть каковы соотношения типов решений и категорий решений по делам с участием конкретного адвоката.

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
8ИМ4А	Паршина Дарья Михайловна

Институт	кибернетики	Кафедра	ВТ
Уровень образования	магистр	Направление/специальность	ИСиТ

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Работа с информацией, представленной в российских и иностранных научных публикациях, аналитических материалах, статистических бюллетенях и изданиях, нормативно-правовых документов.
2. Нормы и нормативы расходования ресурсов	
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого и инновационного потенциала НТИ	Проведение предпроектного анализа. Определение целевого рынка и проведение его сегментирования. Выполнение SWOT-анализа проекта.
2. Разработка устава научно-технического проекта	Определение целей и ожиданий, требований проекта. Определение заинтересованных сторон и их ожиданий.
3. Планирование процесса управления НТИ: структура и график проведения, бюджет, риски и организация закупок	Составление календарного плана проекта. Определение бюджета НТИ.
4. Определение ресурсной, финансовой, экономической эффективности	Приведение оценки экономической эффективности разработки информационно-аналитической системы.

Перечень графического материала (с точным указанием обязательных чертежей):

1. Доли участия исполнителей проекта в запланированных работах
2. Распределение трудоемкости запланированных работ по исполнителям
3. Линейный график запланированных работ
4. Изменение степени готовности проекта в процессе выполнения запланированных работ
5. Структура цены проекта
6. Оценка научно-технического уровня проекта

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
доцент каф. менеджмента ИСГТ	И.С. Антонова	к.э.н., доцент		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8ИМ4А	Паршина Дарья Михайловна		

Заключение

В ходе выполнения данной работы были выполнены следующие задачи: анализ предметной области и текстов судебных решений, проектирование и реализация базы данных для хранения извлеченных параметров, проектирование и разработка инструментов для извлечения атрибутов судебных дел и наполнения базы данных, OLAP анализ данных.

Изучение предметной области позволило выявить актуальность исследования, определить перечень атрибутов, которые возможно и необходимо извлечь из судебных решений, а также спроектировать структуру базы данных и инструментов для извлечения атрибутов судебных дел и наполнения базы данных. На основе спроектированного решения была выполнена программная реализация инструмента для извлечения атрибутов судебных дел в среде разработки Pycharm на языке программирования Python. И выполнена программная реализация инструмента для наполнения базы данных в среде разработки Visual Studio 2013 на языке программирования C#.

OLAP анализ данных проводился в среде Visual Studio 2013. Проведенный анализ позволил получить статистическую информацию и выявить некоторые закономерности между полом судьи и принятым решением, и наказанием. Так, судьи женского пола более склонны к взысканиям, в то время как судьи мужского пола к арестам. Кроме того, судьи женского пола более склонны к решению «Иск удовлетворен», в то время как судьи мужского пола к «Иск не удовлетворен». Или по уголовным делам первой инстанции по статьям главы «Преступления против общественной безопасности» судьи женского пола, в отличии от судей мужского пола, выносят более жесткие решения. Можно сделать вывод, что пол судьи оказывает влияние на принимаемое решение. Кроме того, были построены срезы, показывающие к каким решениям склонны отдельные судьи. Также были показаны какие типы и категории решений преобладают у отдельных

адвокатов. Представленные выводы подтверждаются срезам, представленными в третьем разделе.

В дальнейшем данное исследование можно развивать, увеличив объем анализируемых судебных решений и анализировать не только судебные дела Томской области, а дела всей Российской Федерации. Кроме того, можно увеличить объем параметров, извлекаемых из судебных решений.

Список публикаций

1. Щукова К. Б. , Хлопонин А. А. , Паршина Д. М. Извлечение и анализ данных о судопроизводстве в г. Томске с помощью технологий OLAP и DATA MINING // Технологии Microsoft в теории и практике программирования: сборник трудов XII Всероссийской научно-практической конференции студентов, аспирантов и молодых ученых, Томск, 25-26 Марта 2015. - Томск: ТПУ, 2015 - С. 105-106.

2. Хлопонин А. А. , Паршина Д. М. Обработка решений судов Томской области и города Томска с помощью технологий OLAP и Data Mining // Технологии Microsoft в теории и практике программирования: сборник трудов XIII Всероссийская научно-практическая конференция студентов, аспирантов и молодых ученых, Томск, 22-23 Марта 2016. - Томск: ТПУ, 2016.

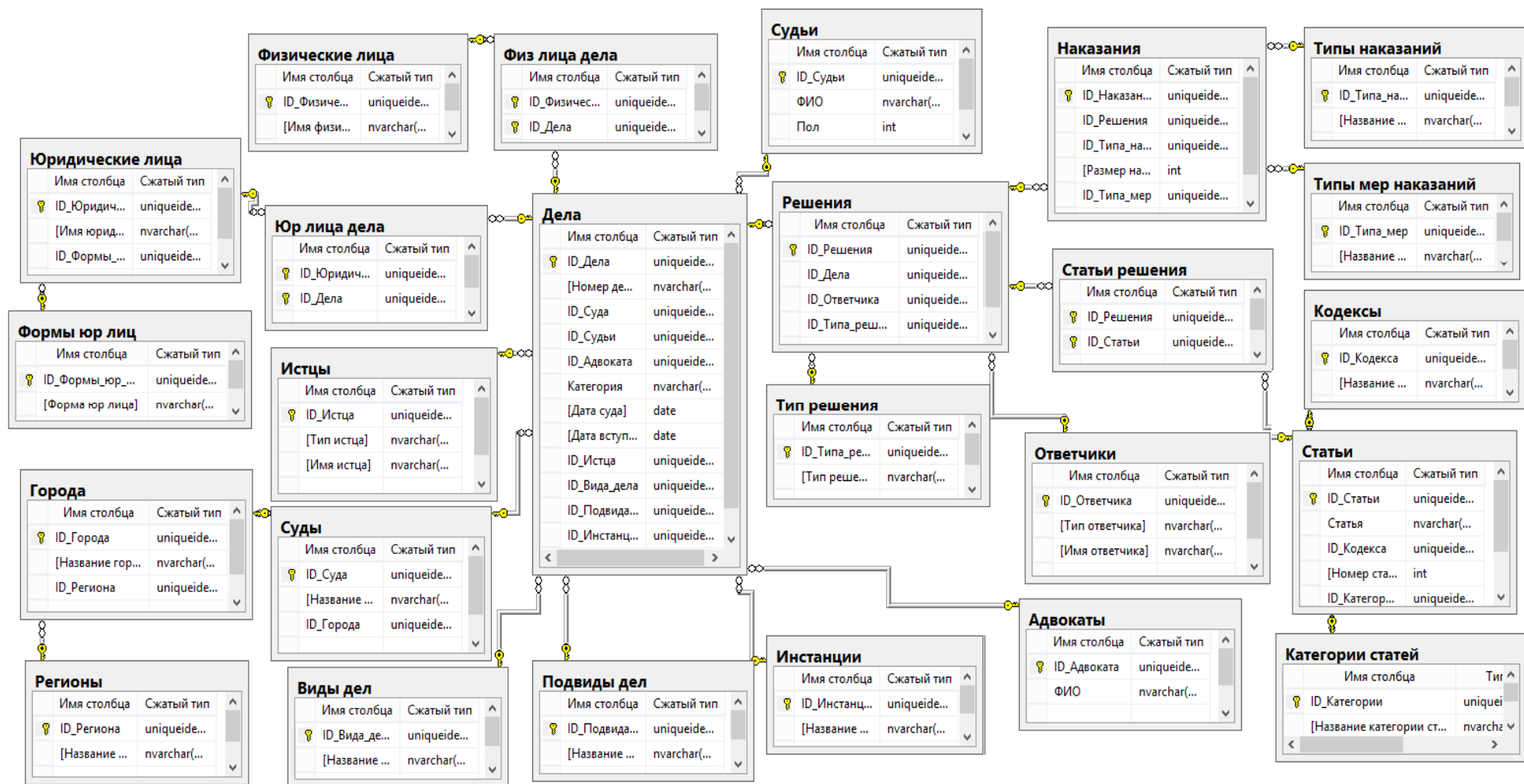
Список использованных источников

3. Федеральный закон N 262-ФЗ от 22.12.2008 (ред. от 05.04.2016) «Об обеспечении доступа к информации о деятельности судов в Российской Федерации».
4. «Гражданский процессуальный кодекс Российской Федерации» от 14.11.2002 N 138-ФЗ (ред. от 02.03.2016).
5. CaseLook [Электронный ресурс]. URL: <http://caselook.ru/> (дата обращения: 28.05.2016).
6. РосПравосудие [Электронный ресурс]. URL: <http://www.rospravosudie.com/> (дата обращения: 28.05.2016).
7. Сравнительный анализ методов определения нечетких дубликатов для WEB-документов / Ю. Г. Зеленков, И. В. Сегалович // Труды 9-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL 2007: Сборник работ участников конкурса. — Переславль-Залесский, Россия. — 2007.
8. Двоичные коды с исправлением выпадений, вставок и замещений символов / В. И. Левенштейн // Доклады Академий Наук СССР — 1965.
9. Регулярные выражения / Фридл, Дж. — СПб., 2001. — 352 с.
10. Журнал «Управление компанией» [Электронный ресурс]. URL: <http://www.zhuk.net/> (дата обращения: 28.04.2016).
11. Введение в OLAP-технологии Microsoft / Наталия Елманова, Алексей Федоров — М., 2002, 272 с.
12. OLAP [Электронный ресурс]. URL: <http://www.olap.ru/> (дата обращения: 28.04.2016).
13. Методы и модели анализа данных: OLAP и Data Mining / Арменак Барсегян, Михаил Куприянов, Валентин Степаненко, Иван Холод — СПб., 2004, 336 с.

14. «Уголовный кодекс Российской Федерации» от 13.06.1996 N 63-ФЗ (ред. от 01.05.2016).
15. СанПиН 2.2.2/2.4.1340 – 03. Санитарно – эпидемиологические правила и нормативы «Гигиенические требования к персональным электронно – вычислительным машинам и организации работы». – М.: Госкомсанэпиднадзор, 2003.
16. СНиП 23-05-95 от 20.05.1995. Строительные нормы и правила Российской Федерации «Естественное и искусственное освещение».
17. СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы: с изменениями от 25 апреля 2007 г. – М.: Информационно-издательский центр Минздрава России, 2003.
18. СанПиН 2.2.1/2.1.1.1200-03 Санитарно-защитные зоны и санитарная классификация предприятий, сооружений и иных объектов. – М.: Госкомсанэпиднадзор России, 2003.
19. Научно-информационный портал Винити. Экология [Электронный ресурс]: – Режим доступа: <http://science.viniti.ru>
20. Технический регламент о требованиях пожарной безопасности: Федеральный закон от 22 июля 2008 года N 123-ФЗ.
21. ГОСТ Р 22.3.03 – 94. Безопасность в чрезвычайных ситуациях. Защита населения.

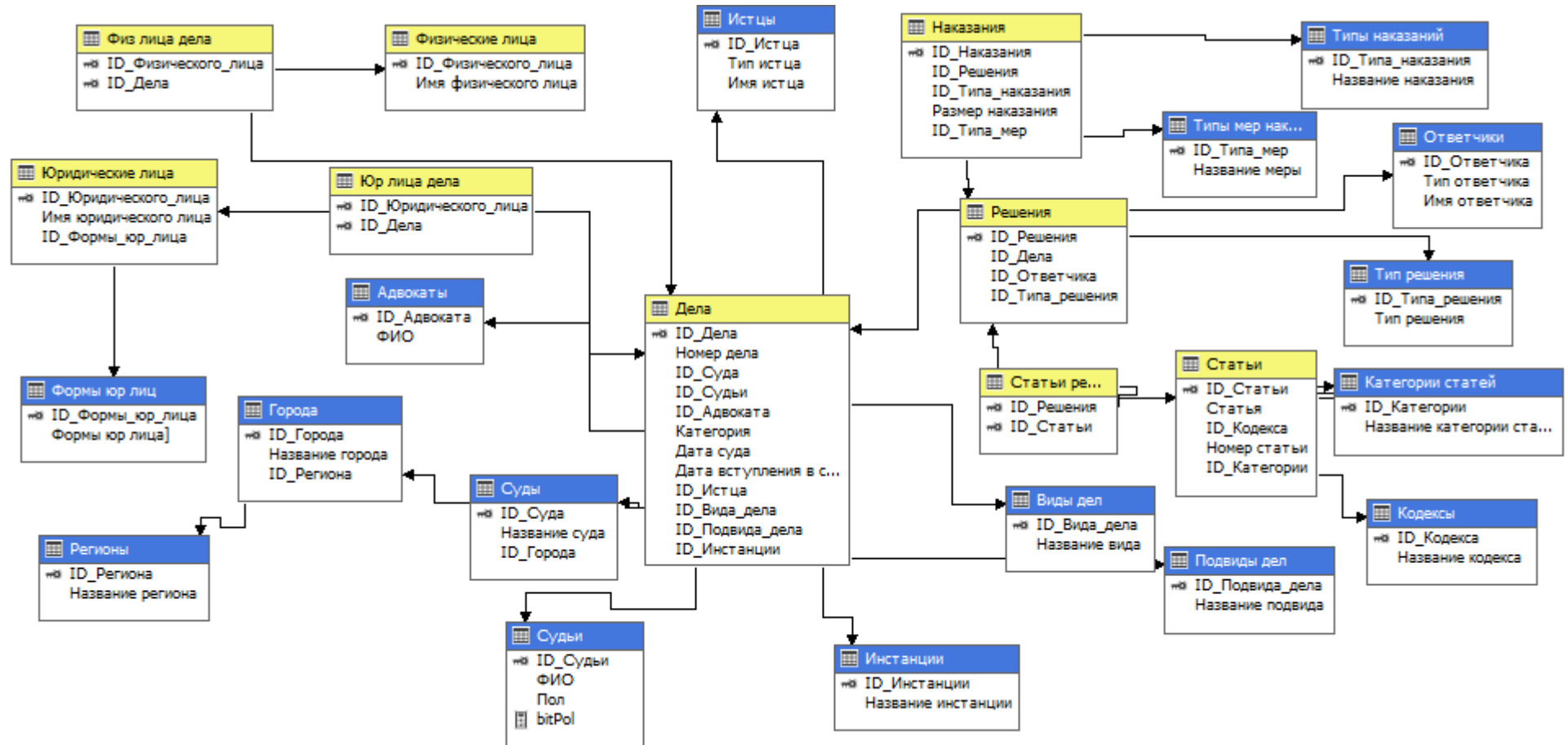
Приложение А

Физическая схема базы данных



Приложение Б

Схема куба данных судопроизводства



Приложение В
Разделы на иностранном языке

Разделы 3, 4
Development of data processing tools.
Data analysis judicial decisions using OLAP technologies.

Студент:

Группа	ФИО	Подпись	Дата
8ИМ4А	Паршина Дарья Михайловна		

Консультант кафедры ВТ:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент каф. ВТ	Мирошниченко Е.А.	к.т.н.		

Консультант – лингвист кафедры иностранных языков ИК:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Старший преподаватель	Шепетовский Д.В.			

3 Development of data processing tools

3.1 Choice of development tools

For extracting of parameters of judicial decisions is used Python programming language in Pycharm development environment. Main advantages of chosen environment and language are:

- readable syntax;
- simple in learning;
- high-level language;
- object oriented language
- wide set of plugins;

For creating the database was used the Microsoft SQL Server 2014 Management Studio environment due to:

- usable graphical interface;
- rich features;
- high performance;
- reliability;

As development environment of application for filling database was used Visual Studio 2013. Choice of this software conditioned with next opportunities:

- work with database objects in SQL Server object explorer;
- tables definition in new constructor of tables;
- development and testing of applications of databases in SQL Server

Express Local DB

- Writing of asynchronous code by simple and intuitive way;
- Obtaining information about the calling object which will help with tracing and debugging;

As programming language was used the C# because it has next advantages:

- object oriented;
- the commitment to safety code;

- versatility;
- relative simplicity.

For creating of OLAP-cubes and execution of OLAP-analysis was used the Visual Studio 2013 environment because it has next advantages:

- Business Intelligence tools;
- working with SQL Server databases;
- useful graphical interace

3.2 Development of the attributes extraction tool from decisions

The tool for extracting attributes from judicial decisions was implemented on the Python programming language with using its different extensions allowing to work on extracting text information.

The main problem of implementation was absence of unified structure and defendants was unnamed. Therefore for some parameters had to make huge list of regular expressions. At that had to complement these lists constantly because decisions in different cases had a large difference. Thus implementation of this tool was iterative.

As a result was obtained the next data: number of case, region, city, kind of case, type of case, instance, name of judge, gender of judge, date of becoming effective, name of standard regulation, article, punishment, the size of punishment, lawyer, defendant, complainant, juridical persons, physical persons.

3.3 Development of the database filling tool

Base on the results of the executed design phase was implemented tool for filling of judicial proceedings database on C# programming language in Visual Studio 2013 environment. As the input data tool takes the collection of XML files.

The main problem in development of this tool was correctly implement relations between different entities and to envisage all possible variants of attributes values. So names can be written by different ways: surname, name,

middle name; surname, initials of name and middle name; initials of name and middle name then surname. Therefore at first name was splitted to separated words then defined the surname and name with middle name. Then it is executes the comparing of surnames by Levenstein distance and separately comparing of name and middle name for coincidence.

As a result of work was get the database filled of judicial decisions attributes. All attributes was written to matched fields of entities.

4 Analysis of a data by OLAP technology

Business Intelligence – is a process of multivariable information extracting and making it into knowledge for effective management of business, executed by the end users with special technologies methods and facilities. One of the BI technologies is the OLAP.

OLAP is an online analytical processing. 12 main rules of OLAP was formulated by E.F. Codd in 1993. He is author of relational databases. Later his definition was recycled to so called FASMI test requiring that OLAP application provides possibility of quickly analisys of separated multidimensional information.

OLAP provides convenient fast means of access, view and analisys of information. User gets native, intuitively understandable model of data organized them as multidimensional cubes. Axis of multidimensional coordinate system are main attributes of analyzed business process. As one of measurements used the time. There are the data quantifically characterized of process on the axis, they called measurements. User who is analyzing information can make the pointcut in different ways, can get summary information (for example by years) or detailed information (by weeks) and make other manipulations, that will cross one's mind during analysis process.

By making use of OLAP it is going to solve next problems:

- collect of statistical information
- checking effect of judges gender on decisions

- checking effect of judges gender on decisions in separate articles and chapters of criminal code of Russian Federation
- checking effect of gender of judges on punishment
- checking of inclination of judges to some decisions or types of judicial cases
- look to statistical of lawyers

Algorithm of making cube:

1. Creating of project in Visual Studio 2013 environment
2. Checking of existing database as data source for further working
3. Tune the view of data source and measurement. Except of main measurements (tables of database) it was added extra measurement of time included of different formats (year, month, day etc.)
4. Creating of cube, and indicating which measurements and measurement groups will be appears

Even 3D cube is difficult for view on computer display to see means of needed measurements not to mention cubes with higher dimensions. For visualize of data kept in cube they usually applied familiar two dimensional, in other words table views which have complex hierarchical headers of strings and columns.

5 Analysis of a data by Data Mining technology

The main phases of Data Mining

1. Data preparing
 2. Indexing of data
 3. Normalize of data for using in handling algorithms
 4. Applying of algorithms and getting results
- 1) Data preparing.

At that phase we must have the data as a result of parsing from sites or documents of courts. That is at this phase assumes presence of text information.

The purpose of this phase is to parse your collected text data, and to store it in a structured format for later use. The goals of the phase are as follows:

1. Visually inspect text files.
2. Parse the files.
3. Serialize complex objects with JSON

Visual inspection of the text documents is an important part of the process of analyzing. During the visual inspection it is identifies parts of documents, in particular there are necessary text parts in judicial decisions, there are parts of text which not necessary takes place as different modifications or completely absent in some judicial decisions. For example, each decision has an information about justice, date of pronouncement of judgment etc. Also, each decision is separated into the text of justification of decision and text of sentence. Besides, each decision has sequence of specific tokens and parts of text that gives possibility for applying different algorithms in handling of text of decisions.

Parsing of text files is an analysis of text using multivariate analyzer with next separation single information entities and complicated object-model of judicial decisions in order to improve its capabilities. For example there is an entire text in beginning of information handling, and after parsing and enrichment of data the object has extra fields and it gives extra opportunities of learning of judicial decisions, opportunity of grouping, sorting and filtering in dataset, decision can have any quantity of fields. That is parsing allows to allocate maximally possible dataset of entities for managing of judicial decisions.

Saving of complexed JSON objects is a moving prepared data into store for next working with them. The database is a NoSQL by principle it is not relational. Such method of data storing provides simplicity of data extracting, exclude the need for complex queries for getting whole object of judicial decision, lets to get simple statistical values on the fly.

2) Indexing of data

Indexing of data can be separated on a few parts: splitting into words, filtering and extraction of stopwords, stemming or lemmatization, indexing itself.

Splitting into words is a procedure of text fragmentation (in this work it is a string splitting of judicial decision) to obtain the data array which has single

element as word, because all words are splitted by space and accordingly fragmentation of string is performed in accordance with by space symbol.

Filtering of words and extraction of stop words is phase with removing of not important words which not affect the analysis of text in future.

Stopwords are words that added subtext to meaning, the extra words which not contain meaning separately and they are auxiliary, also these include prepositions, particles, conjunctions.

Stemming and lemmatization of words is a bringing of words to a normal form. These two methods are different but have similar goal. Stemming is an algorithm, that detach prefixes, suffixes and endings and gives in result cut form of word that has only base part of word. This method is simpler in implementation but can be wrong because splitting can be incorrect.

Lemmatization of word is an algorithm which goal has bringing of word to normal form singular number and the nominative case. This algorithm is harder in implementation but result is better for next using in future.

Indexing of data is the counter of words in the text, each word encountered one time has index 1 accordingly how many times word found in text, that index it has.

3) For using of data in different algorithms it is necessary to bring the data to right form simple text or words dataset not always can match for operations, primarily it is necessary to have the data processing model and understanding whether dataset match with this model or not. If not this algorithm should be rejected in conjunction to loss of use or it needs to get extra information from current dataset.

4) The execution of algorithm is running of software which has implemented algorithm for computing of dataset and to obtain prepared information in needed format.