

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»**
(Н И У « Б е л Г У »)

ИНСТИТУТ ИНЖЕНЕРНЫХ И ЦИФРОВЫХ ТЕХНОЛОГИЙ
КАФЕДРА ОБЩЕЙ МАТЕМАТИКИ

**МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ПРИЛОЖЕНИЕ ДЛЯ АНАЛИЗА
МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ ДАННЫХ**

Магистерская диссертация
обучающегося по направлению подготовки 01.04.01 Математика
очной формы обучения, группы 11120003
Гончарова Дмитрия Викторовича

Научный руководитель
д.т.н., профессор
Аверин Г.В.

Рецензент
Гендиректор
ООО " НПФ МИДИЭЛ"
Гончаров В.А.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1 АНАЛИЗ СОСТОЯНИЯ ВОПРОСА	6
1.1 Анализ существующих баз данных для комплексной оценки многомерных статистических данных	6
1.2 Обзор методов и индексов комплексной оценки	10
1.3 Сложные события при оценке состояния и развития городов и регионов	14
1.4 Обзор существующих Web-ресурсов для оценки данных	16
1.5 Глобальная структура данных	18
2 ВЫБОР И ОБОСНОВАНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ АНАЛИЗА МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ ДАННЫХ.....	21
2.1 Характеристика используемой статистической информации.....	21
2.2 Кластерный анализ статистических данных.....	25
2.3 Обоснование математических моделей для анализа многомерных статистических данных.....	29
3 ПРОЕКТИРОВАНИЕ WEB-ПРИЛОЖЕНИЯ ДЛЯ АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ.....	39
3.1 Выбор технологий и средств реализации.....	39
3.2 Определение функций и задач модулей системы	42

3.3	Разработка структуры приложения и методы реализации	45
3.4	Реализация приложения.	47
ЗАКЛЮЧЕНИЕ		50
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ		52
ПРИЛОЖЕНИЕ А		505

ВВЕДЕНИЕ

Проблема комплексной оценки развития городов, регионов и стран является одной из фундаментальных проблем современной науки. Это направление исследования в своей базовой методологии опирается преимущественно на экспертные методы, которые повсеместно используются в научных и практических целях. Все экспертные оценки являются субъективными, поэтому достоверность их зависит от опыта эксперта и его аналитических возможностей при оценке развития ситуаций. Для того чтобы решить проблему комплексной оценки систем по множеству показателей необходимо уметь моделировать в совокупности природные, экологические и антропогенные процессы. Ориентироваться при этом нужно на вероятностные подходы анализа и описания количественных данных, которые получены в результате статистических наблюдений.

Актуальность данной работы связана с созданием комплексных моделей систем на основе различных методов статистического анализа данных.

Разработкой методов комплексной оценки и анализа рисков занимались Дж.Форрестер, Д.Х. Медоуз, П. Анохин, Р. Уиттекер, А.А. Акаев, М.З. Згуровский, В.Д. Могилевский, К.Боулдинг, В.И. Кузнецов, Е.А. Яйли, Ю. Израэль, В. Маршал, С.Л. Авалиани, А.Б. Качинский, А.В. Кисилев и многие другие авторы.

Цель работы – комплексная оценка состояния и развития городов и регионов на основе статистических данных и исследований моделей характеризующих процессы их развития, а также создание существующего Web-ресурса.

Задачи:

- создание базы данных состояния и развития объектов на основе федеральной службы государственной статистики;

- выбор математических моделей для оценки состояния объектов по статистическим данным и комплексная обработка информации по совокупности показателей;

- проведение предварительного статистического анализа данных, и анализ взаимосвязей;

- установление особенностей и закономерностей развития городов и регионов России на основе комплексной оценки.

Объекты исследования – города и регионы России, показатели состояния их развития.

Предмет исследования – математические модели состояния и развития государственных и территориальных образований.

Методы исследования: исследования выполнены путем использования методов интеллектуального анализа данных и математического моделирования, проведения систематизации данных, применения методов комплексной оценки, а также методов и средств статистического анализа данных. Также были проведены вычислительные эксперименты по полученным математическим моделям и выполнено проектирование Web-приложения.

Структура выпускной квалификационной работы состоит из введения, трёх глав, а также заключения и списка используемой литературы.

В первой производится анализ предметной области. Результаты работы сформулированы в 3 главе.

Во второй главе проводится описание математической модели и алгоритмов для анализа.

В третьей главе рассматривается процесс создания веб ресурса.

1 АНАЛИЗ СОСТОЯНИЯ ВОПРОСА

1.1 Анализ существующих баз данных для комплексной оценки многомерных статистических данных

При изучении многомерных статистических данных городов и регионов России использовались следующие ресурсы:

- рейтинговое агентство «Эксперт РА»;
- федеральная служба государственной статистики (Росстат);
- единая межведомственная информационно-статистическая система ЕМИСС;
- центральный банк Российской Федерации.

Рейтинговое агентство «Эксперт РА» – Российское рейтинговое агентство включено в реестр рейтинговых агентств Банка России. Занимается присвоением рейтинга-гос, а также научно-исследовательской и коммуникационной деятельностью.

Рейтинги агентства «Эксперт РА», наряду с рейтингами некоторых других рейтинговых агентств, влияют на прием финансовых организаций (банков, страховщиков, пенсионных фондов) и эмитентов к тем или иным видам деятельности.

Основные направления деятельности:

1. присвоение кредитных рейтингов частным компаниям и банкам (более 500 компаний в рейтинговых списках агентства на 1 июля 2018 года);
2. анализ регионов и муниципалитетов по показателям кредитования;
3. проведение исследований в банковской сфере, рынков лизинга, факторинга, страхования, МФО, НПФ и т.п. (более 50 исследований и отраслевых обзоров публикуется агентством ежегодно);

4. составление рейтингов компаний в разных отраслях (НПФ, страховые компании, банки, лизинговые компании, МФО, инвестиционные компании, управляющие компании и др.);

5. проведение мероприятий деловой направленности.

Таблица 1.1 – Уровни платежеспособности рейтингового агентства

Категория	Определение
1	2
AAA	Рейтинг объекта характеризуется максимальным уровнем платежеспособности / финансовой надежности / финансовой устойчивости. По данным Агентства, самый высокий уровень платежеспособности / финансовой надежности / финансовой устойчивости в национальном масштабе для Российской Федерации.
AA	Высокий уровень платежеспособности / финансовой надежности / финансовой устойчивости по сравнению с другими объектами рейтинга в Российской Федерации, который лишь немного ниже, чем у объектов рейтинга в рейтинговой категории ruAAA.
A	Умеренно высокий уровень платежеспособности / финансовой надежности / финансовой устойчивости по сравнению с другими объектами рейтинга в Российской Федерации, однако, существует некоторая чувствительность к последствиям негативных изменений в экономической ситуации.
BBB	Умеренный уровень платежеспособности / финансовой надежности / финансовой устойчивости по сравнению с другими объектами рейтинга в Российской Федерации, в то время как чувствительность к воздействию негативных изменений в экономической ситуации выше, чем у объектов рейтинга в рейтинговой категории ruA.

Продолжение таблицы 1.1

1	2
ВВ	Умеренно низкий уровень кредитоспособности/финансовой надежности/финансовой устойчивости по сравнению с другими объектами рейтинга в Российской Федерации. Присутствует высокая чувствительность к воздействию негативных изменений экономической конъюнктуры.
В	Низкий уровень платежеспособности / финансовой надежности / финансовой устойчивости по сравнению с другими объектами рейтинга в Российской Федерации. В настоящее время все еще существует возможность выполнения финансовых обязательств в срок и в полном объеме, но запас прочности ограничен. Способность выполнять обязательства уязвима в случае ухудшения экономических условий.
ССС	Очень низкий уровень платежеспособности / финансовой надежности / финансовой устойчивости по сравнению с другими объектами рейтинга в Российской Федерации. Существует значительная вероятность того, что целевой объект не будет выполнять свои финансовые обязательства в краткосрочной перспективе.
СС	Очень низкий уровень платежеспособности / финансовой надежности / финансовой устойчивости по сравнению с другими объектами рейтинга в Российской Федерации. Существует повышенная вероятность того, что целевой объект не будет выполнять свои финансовые обязательства в краткосрочной перспективе.
С	Очень низкий уровень платежеспособности / финансовой надежности / финансовой устойчивости по сравнению с другими объектами рейтинга в Российской Федерации. Существует очень высокая вероятность того, что целевой объект не выполнит свои финансовые обязательства в краткосрочной перспективе. Своевременное выполнение финансовых обязательств крайне маловероятно.

Федеральная служба государственной статистики (Росстат) - Российский федеральный орган исполнительной власти, занимающийся составлением статистических данных, надзора, связанную со статистической деятельностью на территории РФ.

Основными функциями Федеральной службы государственной статистики являются:

- составление и предоставление статистических данных;
- разработка и совершенствование научно обоснованной официальной статистической методологии для проведения статистических наблюдений и составления статистических показателей, обеспечивающей соответствие этой методологии международным стандартам;
- разработка и совершенствование системы статистических показателей, характеризующих состояние экономики и социальной сферы;
- сбор статистической информации и формирование на ее основе официальной статистической отчетности;
- контроль за выполнением организациями и гражданами, осуществляющими предпринимательскую деятельность без юридического лица, законодательства Российской Федерации в области государственной статистики;
- разработка информационной системы государственной статистики, обеспечение ее совместимости и взаимодействия с другими государственными информационными системами; обеспечение хранения государственных информационных ресурсов и защиты конфиденциальной и отнесенной к государственной тайне статистической информации;
- выполнение обязательств Российской Федерации, вытекающих из членства в международных организациях и участия в международных договорах; осуществление международного сотрудничества в области статистики.

Единая межведомственная информационно-статистическая система (UMISS - UMISS) является государственной информационной системой, которая обеспечивает доступ к официальной статистической информации через Интернет для государственных органов, органов местного самоуправления и других пользователей UMISS. Кроме того, он предусматривает интеграцию статистических ресурсов, обеспечивая их эффективное использование при принятии управленческих решений и прогнозировании, а также представление официальной статистической информации в государственные информационные системы.

Центральный банк Российской Федерации (Банк России) считает себя правопреемником Государственного банка СССР, основными задачами которого являются защита и обеспечение стабильности российского рубля, развитие и укрепление банковского сектора, а также обеспечение эффективного и бесперебойного функционирования платежной системы. Банк России официально является центральным банком Российской Федерации только с 2002 года.

1.2 Обзор методов и индексов комплексной оценки

В практической деятельности для оценки развития городов и регионов России используются различные индексы и показатели:

- индекс человеческого развития;
- индекс глобальной конкурентоспособности;
- индекс экономической свободы;
- индекс экологической эффективности;
- индексы качества и безопасности жизни;
- индекс нестабильности регионов;
- индекс восприятия коррупции;
- покупательная способность среднедушевых денежных доходов;

- оборот розничной торговли на душу населения;
- просроченная задолженность по кредитам, предоставленным физ. лицам;
- инвестиции в основной капитал на душу населения;
- динамика реальных денежных доходов населения;
- динамика вводов жилья;
- индекс образования;
- индекс качества жизни и многие другие.

В области экологической и промышленной безопасности используются индекс загрязнения воздуха, пороговый индекс массы опасных веществ для объектов высокого риска, общий индекс опасности отдельных компонентов, загрязняющих конкретную биогеохимическую среду (вода, воздух и почвы) и т. Д. Индексы помогают представить полученную информацию в интегрированном виде, который помогает обнаруживать сложные явления или дает однозначные оценки.

Индекс экономического «здоровья» регионов используется для оценки текущего уровня развития российских регионов (карта состояний) и для оценки изменений в региональной экономике (карта динамики). Карта состояний позволяет увидеть дифференциацию регионов по текущему уровню развития. Карта динамики позволяет увидеть дифференциацию регионов по темпам изменений: в каких регионах наблюдается позитивная динамика, а в каких идет ухудшение показателей.

Целесообразность разработки данного индекса обусловлена несколькими причинами. Во-первых, существует значительный лаг (более полутора лет), с которым публикуются данные по ВРП (валовому региональному продукту) российских регионов. Причем показатели ВРП не в полной мере отражают ситуацию в регионе: существенная часть ВРП может перераспределяться в пользу Федерального центра, поэтому высокие

показатели ВРП в том или ином регионе не в полной мере отражаются на экономическом здоровье населения, бизнеса и бюджета. Во-вторых, различные ведомства публикуют достаточно большие массивы данных по многим аспектам функционирования региональных экономик, причем публикуют достаточно оперативно, с небольшими временными лагами. Использование, по крайней мере, части этих массивов данных в виде интегральных показателей, позволяет, на наш взгляд, более полно отслеживать как текущее состояние экономического «здоровья» региона, так и направление его изменения.

В качестве источников информации выступают оперативные данные ФНС РФ, Министерства финансов РФ, Центрального Банка РФ и Росстата. При расчете большинства используемых показателей используются данные за последние 12 месяцев (принцип «скользящего окна»), а значения балансовых показателей – на последнюю доступную дату.

Методики расчета интегральных индексов основываются на экспертных методах.

Методики расчетов индекса развития человеческого потенциала (ИРЧП), индекса оценки нищеты населения (ИНН), индекса развития оценки экологического следа и многие другие методики, могут использовать множество показателей, которые составляют в группы.

Методическая оценка индексов чаще проводится на основе нормирования показателей. Например, метод, используемый для вычисления ИРЧП предполагает следующую последовательность вычислений:

Перевод любого показателя x в индикатор, значение которого заключено между 0 и 1 (это позволит складывать различные показатели), осуществляется по формуле:

$$x = \frac{x - \min(x)}{\max(x) - \min(x)}, \quad (1.1)$$

где $\min(x)$ и $\max(x)$ являются минимальным и максимальным значениями показателя x среди всех исследуемых объектов.

Индекс развития человеческого потенциала (ИРЧП) определенной страны зависит от трёх следующих показателей [12 – 13]:

1. Индекс ожидаемой продолжительности жизни $LEI = \frac{LE-20}{85-20}$;
2. Индекс образования $EI = \frac{MYSI+EYSI}{2}$;
3. Индекс средней продолжительности обучения $MYSI = \frac{MYS}{15}$;
4. Индекс ожидаемой продолжительности обучения $EYSI = \frac{EYS}{18}$;
5. Индекс дохода $II = \frac{\ln(GNIpc)-\ln(100)}{\ln(75000)-\ln(100)}$;

ИРЧП является средним геометрическим этих трёх индексов:

$$HDI = \sqrt[3]{LEI \times EI \times II}, \quad (1.2)$$

где LE – ожидаемая продолжительность жизни, MYS – средняя продолжительность обучения населения в годах, EYS – ожидаемая продолжительность обучения населения, еще получающего образование, в годах, $GNIpc$ – ВНД на душу населения по ППС в долларах США.

Аналогичным образом рассчитываются индексы нищеты населения (ИНН-1, ИНН-2).

Все вышеуказанные индексы сформулированы экспертами, исходя из общепринятого соглашения между специалистами. Некоторые индексы и индикаторы, используемые при анализе развития стран, приведены в таблице 1.1.

1.3 Сложные события при оценке состояния и развития городов и регионов России

Изучим сложные события, связанные с наблюдением значений статистических показателей. При статистическом наблюдении в момент записи соответствующих показателей реализуется событие, связанное с определением значимого показателя, как характеристической величины данного события. Наблюдения осуществляются применительно к различным городам и регионам и все последовательности наблюдений представляются временными рядами событий, каждое из которых относится к определенному объекту. Все события являются несовместными, т.к. привязаны к различным годам.

Для случайной величины вероятность события, что в определенный момент времени наблюдаемая величина C меньше некоторого заданного значения c , определяется из функции распределения $P(c) = P(C < c)$, которая находится по данным наблюдений. Эта функция распределения для многих статистических показателей чаще всего подчиняется логарифмически-нормальному закону распределения.

Таблица 1.1 – Некоторые показатели и индикаторы для комплексной оценки городов и регионов

№	Показатель или индикатор	Рекомендуемое значение
Социально-экономическая сфера		
1	Численность населения, тыс. чел.	опр. спецификой задачи
2	Территория, тыс. км ²	-
3	Валовой внутренний продукт на душу населения, \$	max
4	Экспорт товаров, млрд. \$	опр. спец-ой задачи
5	Импорт товаров, млрд. \$	опр. спец-ой задачи
6	Уровень безработицы, %	min
7	Уровень инфляции (потребительские цены)	min
Промышленность и энергетика		
8	Общее потребление энергии (в тоннах нефтяного эквивалента на душу населения), т.н.э./чел.	опр. спец-ой задачи
9	Потребление электроэнергии, ГВт в час	опр. спец-ой задачи

Для одного объекта все подобные наблюдаемые события (за год, месяц, сутки и т.д.) образуют полную группу. Для дискретных событий $\sum_{i=1}^k P_i = 1$, для непрерывных случайных величин $\int_0^{\infty} f(c)dc$, где i – текущий номер наблюдения событий в разные моменты времени, k – количество наблюдений, а $f(c)$ – функция распределения случайной величины, причем $c \geq 0$.

Если происходит наблюдение одной величины на нескольких точках наблюдения (в пределах одного региона и города), то все события также образуют полную группу или их вероятности могут быть нормируемы. Для дискретных событий $\sum_{i=1}^k P_i = 1$, для непрерывных случайных величин $\int_0^{\infty} f(c)dc$, где i – номер объекта наблюдений, c – среднегодовое значение наблюдаемой величины, m – количество объектов наблюдения, а $f(c)$ –

функция распределения случайной величины, которая оценивается по данным для всех объектов. Все события в указанном случае можно рассматривать как совместные, так как они привязываются к одному и тому же моменту времени – годам.

При событийной оценке нескольких статистических величин на одном объекте (в определенном городе и регионе) реализуется совместное событие одновременного наблюдения указанных величин.

Для данного совместного события, в случае если все события независимы, вероятность сложного события равна произведению вероятностей более простых событий:

$$P_n(c) = P(c_1) * P(c_2) * ... * P(c_n), \quad (1.3)$$

где c_1, \dots, c_n – количественное значение показателей для различных природных ископаемых.

Для зависимых событий вероятность сложного события будет равна

$$P_n(c) = P(c_1) * P_{c_1}(c_2) * ... * P_{c_{n-1}}(c_n), \quad (1.4)$$

где условные вероятности $P_{c_{i-1}}(c_i)$ вычисляются в предположении, что все предыдущие события, связанные с добычей полезных ископаемых, произошли. Вероятности таких событий могут быть оценены по опытным данным путем определения распределений вероятности сложных и более простых событий и изучения взаимосвязей между ними.

1.4 Обзор существующих Web-ресурсов для оценки данных

Data and Analytics создает инструменты, которые улучшают качество исследований по оценке воздействия на основе анализа данных. Мы пользуемся преимуществом исследований в DIME для разработки и тестирования решений, во-первых, для обеспечения высокого качества сбора данных и качества исследований по всему профилю, и, во-вторых, для

обеспечения открытого обучения и инструментов для более широкого сообщества разработчиков. Все программные инструменты, руководства по исследованиям и тренинги, которые разрабатывает команда аналитиков Data and Analytics, общедоступны.

Анализ принятия решений (модель DIME), принятые решения обычно сосредоточены вокруг полезности государственных инструментов власти. Модель DIME - это традиционная абстрактная категоризация действий, основанная на аспектах национальной власти. Каждая категоризация - дипломатическая, информационная, военная и экономическая - является инструментом национальной власти (IOP).

Главными публичными ресурсами для анализа являются:

DIME Wiki

DIME wiki - это универсальный магазин ресурсов на всех этапах оценки воздействия: дизайн, полевые работы, данные и анализ. Каждая статья содержит краткое изложение лучших практик и ключевых ресурсов для успешного выполнения конкретной задачи оценки воздействия. Весь контент DIME Wiki общедоступен и доступен для редактирования, а группа DIME Analytics активно привлекает других лидеров в этой области.

ietoolkit

Флагманский программный пакет DIME Analytics, ietoolkit, представляет собой набор Stata, который рутинно распределяет общие аналитические задачи при оценке воздействия. Его можно установить через SSC, а исходный код доступен для публичного просмотра и публикации на GitHub. Спрос на ietoolkit очевиден: программный пакет загружался в среднем 304 раза в месяц в течение трех месяцев после его запуска в блоге Development Impact. В марте 2018 года он занял 3-е место в списке хранилищ Stata GitHub.

Репозитории GitHub группы Всемирного банка

GitHub - это инструмент контроля версий, который позволяет пользователям обмениваться кодом и совместно работать над ним. Группа Всемирного банка GitHub имеет два основных типа репозитория: ресурсы воспроизводимости исследований для отдельных проектов (например, «Вода, когда она закончится») и обобщенные репозитории кода с инструментами, которые полезны для разных проектов (например, код Stata). Эти репозитории позволяют исследователям обнародовать свой код и результаты исследований для прозрачности и в качестве ресурса для сообщества разработчиков.

Управление успешными оценками воздействия

Флагманское обучение — DIME Analytics - это недельный ежегодный курс, открытый для общественности. Это практическое обучение предназначено для улучшения навыков и знаний разработчиков оценки воздействия, ознакомления их с критическими проблемами в реализации ИЕ, повторяющимися проблемами и передовыми технологиями. Места для личного обучения открыты для общественности, причем приоритет отдается кандидатам из правительств развивающихся стран, исследовательских институтов и некоммерческих организаций. Удаленный доступ к обучению через WebEx открыт для всех. Все презентации и учебные материалы курса (включая лабораторные упражнения и решения) общедоступны через Open Science Framework.

1.5 Глобальная структура данных

Глобальная структура данных (Global Data Structure Definition) - это Инициатива по обмену статистическими данными и метаданными (SDMX) направлена на установление статистических и технических стандартов для облегчения обмена статистическими данными и метаданными с использованием современных информационных технологий¹. SDMX - это

стандарт ISO, разработанный для описания статистических данных и метаданных, нормализации их обмена и повышения эффективности их обмена статистическими и аналогичными организациями. Ряд государств-членов и организаций участвуют в разработке этих стандартов для различных областей официальной статистики. SDMX состоит из:

- технических стандартов (включая информационную модель),
- статистических указаний и • ИТ-архитектуры и инструментов.

Однако SDMX - это больше, чем формат для обмена данными. Вместе технические стандарты, статистические рекомендации, а также архитектура и инструменты ИТ могут поддерживать улучшенные бизнес-процессы для любой статистической организации.

Определение структуры данных (DSD) является центральным компонентом SDMX. Он определяет:

1. Набор концепций (концептуальная схема), которые описывают и идентифицируют набор данных. Эти понятия бывают трех типов; измерения - это понятия, которые идентифицируют и описывают данные (например, период времени), атрибуты - это понятия, которые только описывают данные (например, торговая система), а измерения представляют ценность конкретного наблюдения. В наборе данных комбинация измерений должна быть уникальной, но не для атрибутов.

2. Уровень вложения как измерения, так и атрибута (на уровне набора данных, группы, серии / раздела, наблюдения), а также их статус (обязательный или условный).

3. Списки кодов, которые обеспечивают приемлемые значения для концепций в их предлагаемом стандартном кодировании и форматировании. Определение структуры данных может использоваться для описания данных временных рядов, данных поперечных и многомерных таблиц.

Согласованное и принятое в глобальном масштабе DSD в любой области официальной статистики способствует обмену статистическими

данными с использованием современных информационных технологий, сводя к минимуму необходимость вмешательства человека. С учетом этого утверждена межведомственная рабочая группа в составе Евростата, Международного торгового центра (МТЦ), Организации экономического сотрудничества и развития (ОЭСР), Статистического отдела Организации Объединенных Наций (СОООН) и Конференции Организации Объединенных Наций.

Торговля и развитие (UNCTAD) подготовили предлагаемый DSD для Международной статистики торговли товарами (IMTS). В рамках общего DSD были предложены три различных DSD, каждый из которых использует различное подмножество основных концепций DSD для следующих целей:

- 1) сбор данных,
- 2) публичное распространение
- 3) совокупное распространение.

Глобальные DSD для других статистических По мере возможности DSD IMTS повторно использует концепции и списки кодов, которые были согласованы на международном уровне. Рабочая группа инициировала рассмотрение предлагаемого Глобального DSD для IMTS. Что касается содержимого IMTS-DSD, обзор должен охватывать концепции (как описано ранее) и соответствующие списки кодов.

2 ВЫБОР И ОБОСНОВАНИЕ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ АНАЛИЗА МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ ДАННЫХ

2.1 Характеристика используемой статистической информации

Статистические данные для анализа были использованы и взяты с официального сайта Федеральной службы государственной статистики России (Росстат) [1]. Данная служба является федеральным органом исполнительной власти, осуществляющим функции по сбору и обработке официальной статистической информации о социальном, экономическом, демографическом и экологическом положении, а также функции по контролю и надзору в области государственной статистической деятельности.

На основе имеющихся данных была сформирована база данных социально-экономических показателей регионов и городов России.

Соответствующая база данных, характеризующая состояние регионов и городов, включает информацию по каждому из 80 регионов и по основным из 150 городам страны.

Для каждого региона использовалась информация по 8 социально-экономическим показателям:

- объем товаров и выполненных работ, и услуг собственными силами. Добыча полезных ископаемых, млн. руб.;
- объем товаров и выполненных работ, и услуг собственными силами. Обрабатывающие производства, млн. руб.;
- объем товаров и выполненных работ, и услуг собственными силами. Производство и распределение электроэнергии, газа и воды, млн. руб.;
- продукция сельского хозяйства, млн. руб.;
- объем работ, выполненных в строительстве, млн. руб.;

- объем платных услуг населению, млн. руб.;
- оборот розничной торговли, млн. руб.

Ещё использовались показатели, характеризующие состояние и развитие регионов. Данные формировались с шагом два года в период с 2003 по 2017 годы. Вид данных в виде таблицы Excel приведен на рисунке 2.1.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Регионы	Выпуск товаров и услуг в момент т, млн. руб., 2015	Добыча полезных ископаемых в момент т, млн. руб., 2015	Обрабатывающие производства в момент т, млн. руб., 2015	Производство и распределение электроэнергии, газа и воды в момент т, млн. руб., 2015	Продукция сельского хозяйства в момент т, млн. руб., 2015	Объем работ, выполненных по виду экономической деятельности «Строительство» в момент т, млн. руб., 2015	Объем платных услуг населению в момент т, млн. руб., 2015	Оборот розничной торговли в момент т, млн. руб., 2015	Производство сельскохозяйственного хозяйства в момент т, млн. руб., 2015	«Строительство» в момент т, млн. руб., 2015	Объем платных услуг населению в момент т, млн. руб., 2015	Оборот розничной торговли в момент т, млн. руб., 2015	Выпуск товаров и услуг в момент т, млн. руб., 2012	Добыча полезных ископаемых в момент т, млн. руб., 2012	Обрабатывающие производства в момент т, млн. руб., 2012	Производство и распределение электроэнергии, газа и воды в момент т, млн. руб., 2012	Продукция сельского хозяйства в момент т, млн. руб., 2012	Объем работ, выполненных по виду экономической деятельности «Строительство» в момент т, млн. руб., 2012	Объем платных услуг населению в момент т, млн. руб., 2012	Оборот розничной торговли в момент т, млн. руб., 2012	Производство сельскохозяйственного хозяйства в момент т, млн. руб., 2012
2	1052128	11810786	1491884	1E+05	1696398,3	2692079	9278379,5	263021	299631,4	574678	1774813,2										
3	81211	562897	28740	220692	61981,2	74393	275878,4	23903	7383,4	7776	29081,0										
4	425	171551	15549	73896	24563,8	45074	217652,8	13897	2707,4	8374	24600,7										
5	3406	376168	33221	37257	33119,3	46909	195264,6	12128	3226,8	7820	20234,0										
6	9250	236550	59744	201094	73114,4	110391	463696,3	29419	8210,5	12755	52608,7										
7	933	114727	28550	17296	21708,9	38453	148132,7	6179	2245,5	6077	14123,6										
8	4582	456942	20632	37367	51533,9	43478	168897,4	9395	4375,5	6806	23072,7										
9	725	108058	34179	22817	9805,5	24043	85533,8	7971	2901,3	3151	12613,8										
10	44940	151800	59216	122168	44569,9	48392	179787,8	19113	4372,5	5968	23239,8										
11	9202	926869	28941	101250	40965,5	65228	215896,2	16552	5372,7	6888	26135,5										
12	15495	2086349	268205	103624	320716,2	452056	1725247,4	35958	41590	67088	194375,0										
13	147	104302	12744	64624	19207	30770	114912,1	13289	3352,1	4747	17895,9										
14	2402	250368	32930	54441	26500,9	39834	168427,8	12211	6488,8	6434	23991,4										
15	1026	178438	15431	24271	25570,1	31238	157362,9	9319	6501,2	5601	26652,8										
16	173	126629	15502	126701	32457,4	44634	192429,3	16953	3511,5	6774	23828,3										
17	1266	198376	88438	32380	21202,5	44361	211105,3	11111	9963,5	8027	28151,7										
18	5650	524308	48251	56158	31261,1	64389	253383,0	15284	5626,6	9645	29490,2										
19	1224	310374	32714	33950	38632,1	48324	204565,0	10438	7661,8	9701	25567,7										
20	878269	5158077	628697	9811	821597,5	1428114	4910106,7	0	174140,3	391045	1178970,3										
21	739111	4879272	522267	245104	807347	816165	2609919,3	68571	126037	247688	411326,6										
22	46286	76439	24764	6137	12658,9	28843	103880,9	2558	3812,4	5630	19853,7										

Рис. 2.1 – Пример представления таблицы данных по регионам России

При изучении уровня экономического развития городов в определенный год (в период 2003 – 2017 гг.) использовались следующие основные показатели для 153 городов России:

- объем товаров, выполненных работ и услуг собственными силами. Обрабатывающие производства, млн. руб.;
- объем выполненных работ в строительстве, млн. руб.;
- оборот розничной торговли, млн. руб.;
- объем инвестиций в основной капитал, млн. руб.

Кроме перечисленных использовались еще множество дополнительных основных показателей, характеризующих состояние и развитие городов.

Данные формировались с шагом два года в период с 2003 по 2017 годы. Вид данных в виде таблицы Excel приведен на рисунке 2.2.

Классификация всех используемых данных по регионам и городам России приведена в таблице 2.1. Из данных показателей формировались таблицы данных, где к строкам привязывались регионы или города, а в столбцах находились соответствующие значения показателей, которые, в свою очередь, группировались по времени наблюдений.

	A	B	C	D	E	F	G	H	I	J
1		Численность населения (оценка на конец года), тыс. человек	инвестиции в основной капитал, млн. руб.	объем работ, выполненных по виду деятельности «Строительство»	Оборот розничной торговли, млн. руб.	Среднегодовая численность работников организаций, тыс. человек	Среднемесячная номинальная начисленная заработная плата, руб.	Инвестиции в основной капитал, млн. руб.	Наличие основных фондов организаций (на конец)	Ввод в действие основных фондов, млн. руб.
3	Белгород	387,1	61540,5	7303,1	39758,3	113,6	30011,1	18783,5	316414,5	22652,1
4	Брянск	424,6	69661,8	4026,1	49842,5	130,2	26839,3	17911,5	279434,9	20808,8
5	Владимир	357,4	63481,3	1266,7	38949,6	110,7	29349,7	14785,8	253917,2	12948,1
6	Воронеж	1032,4	145436,5	15910	104824	277,1	30819,6	63260,8	71399,3	60459,7
7	Иваново	408	26500,6	1697,5	31242,1	91,6	26286,7	6916,7	203928,4	7260,6
8	Калуга	342	195420,4	21581,3	41236,8	119,3	35318	42402,6	408448,1	46592,9
9	Кострома	276,7	36124,4	229	12015,2	67,2	26807,4	5278	128385,7	5889,6
10	Курск	443,2	68872	10455	40534,8	119,2	26409,3	19420	202076	12061,5
11	Липецк	510	397155,1	8575,4	51936,4	146,7	31764,7	30782,6	569189,6	569189,6
12	Орел	319,7	43281,3	4364,2	31720,4	93,7	26738,8	16019,5	208912,4	18172,1
13	Рязань	534,8	164999,6	4949,2	57786	160,8	31033,1	26832,8	353321,6	32518,2
14	Смоленск	328,9	51098	2307,6	55565,1	106,9	28939	6069,9	150340,5	7216,3
15	Тамбов	288,4	47348,5	4979,4	38221,1	93,2	26289,7	20462	247306,1	11453,4
16	Тверь	416,4	62824,6	2357,5	51520,6	114,6	31659	15002	26896,3	16956
17	Тула	551,6	233287,6	5977,6	59632,2	175,4	33804	28307,6	354709	27972,7
18	Ярославль	606,7	150492,9	5698,1	52055,8	175,1	32838,9	31127	458497,7	19523,2
19	Москва	12330,1	4789573	821588	4310107	4662,1	64310	1611512	12243251	1510365,7
20	Старый Оскол	222,1	134504	3692,7	19544,5	68,3	29502,4	...	115042,2	6614,5
21	Обнинск	111,4	38351,8	709,3	9495,7	29,8	40389	5053,8	47558,8	5378,3
22	Елец	105,4	11520,8	959,8	7525,1	26,8	23093,7	4855,4	48281,8	48281,8
23	Балашиха	428,4	41998,2	4016,3	52673,9	29,8	47367,5	5641,8	85310,9	19441,6
24	Железнодорожный									
25	Жуковский	108,4	10704,1	309	5630,8	25,4	51506,6	3142	33834,4	3175
26	Коломна	143,6	30967,9	1438,3	12808,4	41,4	38080,2	5506,3	74811,2	6528,3
27	Королев	221,1	45783,5	3075,9	27043,2	54,8	52237,6	7151,8	112321,3	18844,4
28	Люберцы	190,5	11451,5	527,2	12928,3	25,4	49697,1	11835,9	64459,2	11879,3
29	Мытищи	201,1	60409,8	4426,2	120393,6	53,9	47652,3	28439,1	138190,9	22822,1
30	Ногинск	101,8	28199,3	76,9	21920,8	29,5	47313,8	12135,8	87459,8	14183,1
31	Одинцово	141,3	22893,4	11934,7	6834	29,6	44220,8	4137,2	48024,2	7454,7
32	Орехово-Зуево	120,2	13335,9	853,9	14691,1	20,2	36390	1662,7	21735,2	3407,1
33	Полдьск	292,8	48407,5	5674,9	18482,5	58,5	49794,5	8998,2	113450,6	14316,9

Рис. 2.2 – Пример представления таблицы данных по городам России

В дальнейшем полученные таблицы данных обрабатывались методами статистического анализа с использованием программного продукта Statistica

для установления зависимостей характеризующих социально-экономического развитие.

Таблица 2.1. – Классификация данных, характеризующих состояние и развитие регионов и городов России

№ п/п	Группы показателей	Количество показателей
Регионы России		
1	Социально-экономические показатели	10
2	Население	3
3	Занятость и безработица	3
4	Уровень жизни населения	5
5	Валовый национальный продукт	2
6	Предприятия и организации	3
7	Добыча ископаемых, обрабатывающие производства, производство и распределение энергии, газа и воды	4
8	Сельское хозяйство	3
9	Строительство	3
10	Инвестиции	2
11	Внешнеэкономическая деятельность	3
Всего		41
Города России		
1	Социально-экономические показатели	6
2	Инфраструктура	3
3	Население	3
4	Занятость и заработная плата	5
5	Социальная поддержка населения	3
7	Добыча ископаемых, обрабатывающие производства, производство и распределение энергии, газа и воды	4
8	Охрана окружающей среды	4
9	Строительство жилья	2
15	Деятельность предприятий	4
16	Розничная торговля и общественное питание	4
20	Фонды и инвестиции	2
21	Финансовая деятельность	4
Всего		50

Общий объем введенной и используемой для анализа статистической информации составил более 10 тысяч данных наблюдений. Сформированные массивы данных относятся к многомерным статистическим данным.

2.2 Кластерный анализ статистических данных

Анализ информации о состоянии и развитии стран регионов и городов России основывался на данных Федеральной службы статистики. Соответствующая база данных, характеризующая состояние социально-экономических показателей регионов и городов подгружалась в разрабатываемое Web-приложение. В соответствующем модуле приложения реализованы возможности выбора до 2 – 6 показателей, которые могут выступать в виде переменных состояния объектов при многомерном анализе данных. В этом случае из исходных таблиц данных формировались рабочие таблицы, показатели которых выбирает пользователь.

В Web-приложении предусматривается использование нескольких модулей статистического анализа данных, которые необходимы пользователю при изучении закономерностей в данных. Соответствующие модули перечислены ниже:

- модуль определения основных статистик;
- модуль корреляционного и регрессионного анализа данных;
- модуль кластерного анализа.

Модуль определения основных статистик позволяет найти основные статистики в группе данных: среднее, дисперсию, среднеквадратичное отклонение, максимум, минимум и т.д.

Модуль корреляционного и регрессионного анализа данных дает возможность установить существование связей между переменными и найти простейшие регрессионные уравнения.

Модуль кластерного анализа позволяет провести упорядочивание регионов или городов в сравнительно однородные группы. При выборе относительно большого количества кластеров возможно многопараметрическое ранжирование объектов и установление их рангов.

Для примера выполним кластерный анализ данных регионов России для чего выберем из массива данных пять переменных:

- валовый региональный продукт на душу населения;
- среднедушевые денежные доходы населения;
- средний размер назначенных пенсий;
- объем экспорта, пересчитанный по курсу доллара;
- объем импорта, пересчитанный по курсу доллара.

Изучение особенностей развития регионов России осуществлялось методом кластерного анализа данных для сравнения состояний регионов. Кластеризация проводилась методом k -средних в программе Statistica для указанной группы показателей, характеризующих социально-экономическую устойчивость регионов. Методика кластеризации предполагала применение метода ближайшего соседа, где в качестве меры кластеризации использовалось евклидовое расстояние. В соответствии с методикой анализа изучаемые показатели предварительно нормировались путем приведения их к виду: $z_k^{st} = (z_k - z_k^{sr}) / \sigma_k$, где z_k^{sr} – среднее значение k -того показателя, σ_k – среднеквадратичное отклонение. Число кластеров определялось методом иерархической кластеризации в программном продукте Statistica путем построения дендрограмм (рис. 2.3).

Кластеризация регионов по данным наблюдений значений показателей социально-экономической устойчивости позволила выделить пять групп регионов.

Изменение средних значений стандартизированных показателей для кластеров показано на рисунке 2.4.

Таким образом, первый кластер содержал 7 регионов, второй 11, третий 16, четвёртый 20 и пятый 27 регионов:

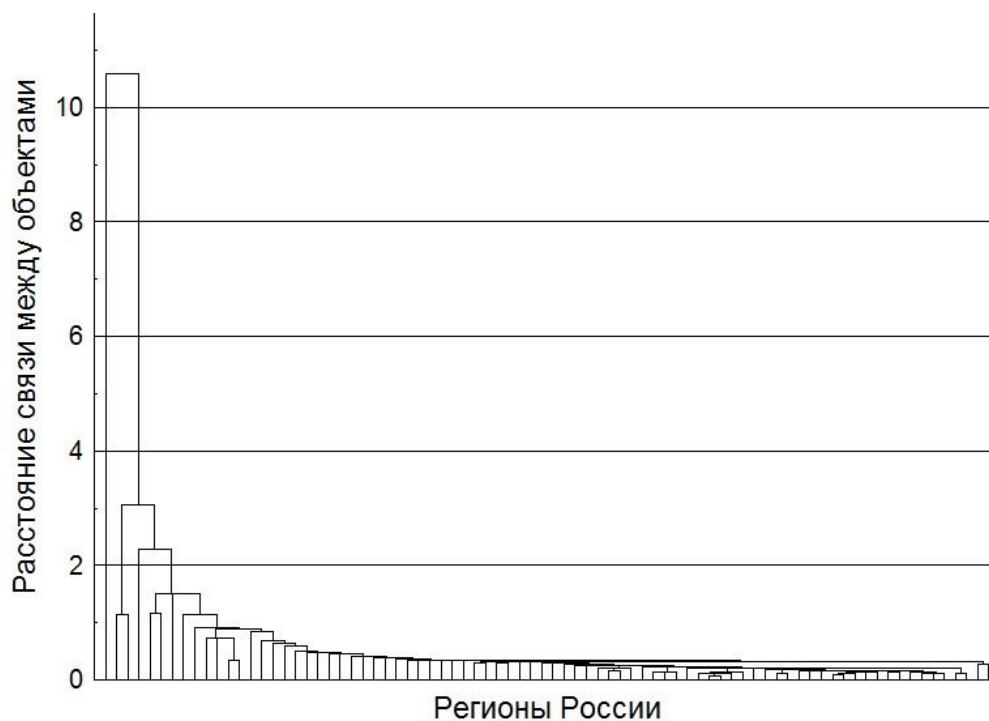


Рис 2.3. – Древоидная диаграмма распределения регионов по кластерам для данных 2015 года и пяти переменных

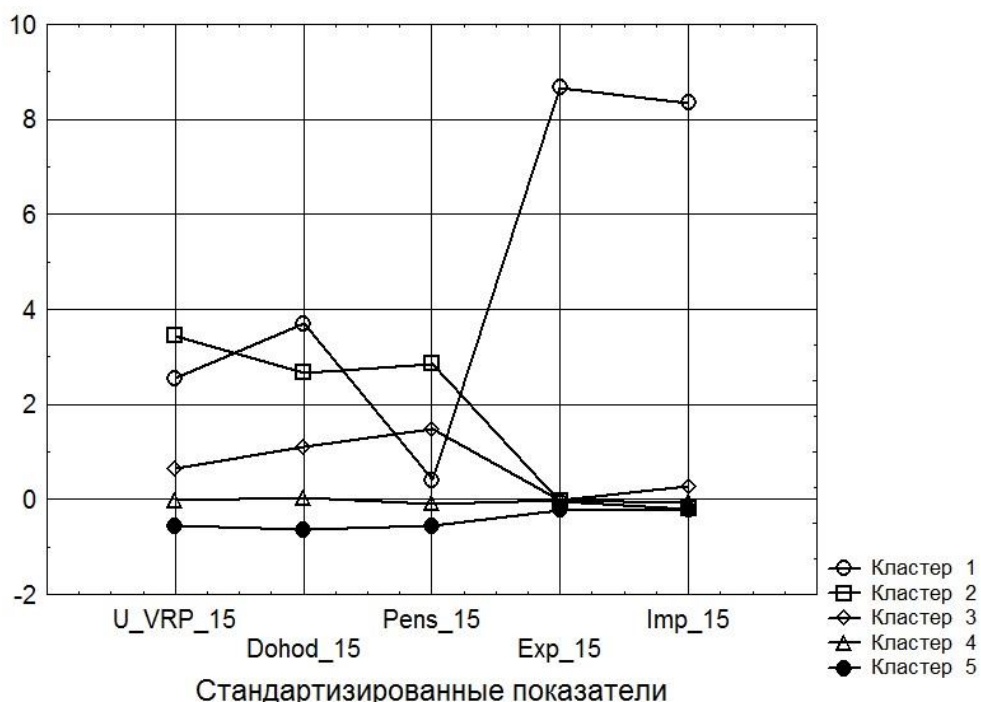


Рис. 2.4. – Средние значения стандартизованных показателей для кластеров 1, 2, 3, 4, 5.

- первый кластер – города Москва и Санкт-Петербург, Тюменская и Томская области, Республика Саха (Якутия), Сахалинская область и Чукотский автономный округ;

- второй кластер – Белгородская, Калужская, Липецкая, Московская области, Республика Карелия, Республика Коми, Архангельская, Вологодская, Калининградская, Ленинградская, Мурманская,

- третий кластер Новгородская области, Республика Татарстан, Пермский край, Самарская, Свердловская области, Республика Алтай, Республика Хакасия, Красноярский край, Иркутская, Кемеровская, Томская области, Камчатский край, Приморский край, Хабаровский край, Амурская область;

- четвертый и пятый кластеры – все оставшиеся регионы России.

Таким образом, выполненный кластерный анализ состояния регионов по нескольким социально-экономическим показателям позволяет провести ранжирование объектов исходя из уровня их развития.

2.3 Обоснование математических моделей для анализа многомерных статистических данных

В настоящее время имеется ряд методик оценки состояния и развития регионов и городов России, в которых используются различные методы анализа социально-экономических данных [2 – 10]. В литературе изложены основные методы моделирования социально-экономических процессов [11 – 22].

Математические модели анализа данных основаны на методах регрессионного анализа.

Считают, что переменные зависимы, если их значения каким-то образом согласованы друг с другом в имеющихся наблюдениях. Реальные процессы или объекты могут характеризоваться набором переменных, которые бывают зависимые и независимые.

Независимые переменные (входные данные, показатели-аргументы, предикторные переменные) - это переменные, описывающие условия формирования изучаемого реального процесса или функционирования объекта. Это переменные, которые можно задавать, измерять, частично контролировать или регулировать.

Зависимые переменные (результат, ответы, результат или объяснение) являются переменными, которые характеризуют процесс или результат (эффективность) функционирования объекта. Обычно это переменные, которые позволяют предсказать процесс или описать объект.

Случайные переменные (скрытые, остаточные) - скрытые, которые не являются непосредственно измеримыми случайными остаточными компонентами, отражающими влияние на зависимые переменные незарегистрированных факторов, а также случайные ошибки при измерении или определении показателей. Часто эти переменные называются «остатками».

Зависимые и независимые переменные при обработке социально-экономических данных обычно являются количественными, то есть скалярно измеряют в определенной шкале определенные свойства (ВРП, денежный доход, количество работников, данные об экспорте и импорте и т. Д.).

Ключевым понятием, описывающим связи между переменными, является корреляция. Две переменные могут быть связаны либо функциональной зависимостью, либо статистической, либо быть независимыми друг от друга.

Статистическая зависимость - это отношение, в котором изменение одной из величин вызывает изменение в распределении другой. Если изменение одного из значений изменяет среднее значение другого значения, то такая статистическая зависимость называется корреляционной.

Коэффициент корреляции – характеризует степень тесноты связи между нормально распределёнными случайными переменными X и Y . Выборочное значение r коэффициента корреляции подсчитывается по формуле:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.1)$$

Значение r является измерителем степени тесноты линейной статистической связи между переменными и изменяется в пределах $-1 \leq r \leq 1$. При $|r|=1$ соотношение (2.1) подтверждает чисто функциональную линейную зависимость между переменными X и Y , при $r=0$ свидетельствует о полной независимости переменных. Положительные значения коэффициента корреляции указывают на одинаковый характер тенденции взаимосвязанного изменения величин X и Y (например, увеличивается X и увеличивается Y), отрицательные значения указывают на противоположную тенденцию. В случае если распределения величин X и Y отличаются от нормального или одна из величин не является случайной, коэффициент корреляции можно использовать лишь в качестве одной из возможных характеристик степени тесноты связи.

Оценка значимости коэффициента корреляции основывается на проверке гипотезы об отсутствии корреляционной связи между переменными [23, стр. 327]. Известно, что величина $\bar{t}(n-2)$ при условии малых значений r распределена по закону Стьюдента с $n-2$ степенями свободы. Поэтому если окажется, что

$$\bar{t}(n-2) = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} \leq t_{0,05}(n-2), \quad (2.2)$$

то гипотеза об отсутствии корреляционной связи принимается. Если $\bar{t}(n-2) \geq t_{0,05}(n-2)$, то коэффициент корреляции значимо отличается от нуля, а величины X и Y коррелированы. Здесь $t_{0,05}(n-2)$ – 5%-ная точка распределения Стьюдента с $n-2$ степенями свободы. Следует иметь в виду, что значимость коэффициента корреляции сильно зависит как от его величины, так и объёма выборки по которой он вычислен.

Корреляционное отношение – измеритель степени тесноты корреляционной связи любой формы (в том числе и нелинейной). Для определения корреляционного отношения область значений независимой переменной X разбивают на интервалы группирования, определяют средние ординаты \bar{y}_i для каждого интервала группирования и рассчитывают корреляционное отношение ρ_{xy}

$$\rho_{xy} = \frac{s_{y(x)}^2}{s_y^2}, \quad (2.3)$$

где $s_{y(x)}^2$ – дисперсия средних \bar{y}_i около общего среднего \bar{y} , s_y^2 – общая выборочная дисперсия величины y [23, стр. 270-274]. Значения корреляционного отношения лежат в пределах $0 \leq \rho_{xy} \leq 1$, причем $\rho_{xy} \geq |r|$.

Если исследуется связь между несколькими переменными (более двух), то корреляцию в этом случае называют множественной. Степень тесноты множественной связи оценивается множественным коэффициентом корреляции R . Квадрат величины R называют коэффициентом детерминации. Множественный коэффициент корреляции изменяется в пределах $0 \leq R \leq 1$. Он оценивает, какая доля дисперсии исследуемой зависимой переменной определяется через функцию регрессии кумулятивным эффектом независимых переменных. В статистике в специальных вычислительных процедурах коэффициент множественной корреляции определяется либо матрицей парных коэффициентов корреляции, либо вектором частных коэффициентов корреляции.

Наряду с группированием и визуализацией данных, расчет корреляций является стандартным начальным этапом любого исследования, связанного с анализом данных.

Регрессионный анализ данных позволяет установить зависимости между переменными вида $\bar{y}_x = f(x)$. Условное среднее \bar{y}_x – среднее арифметическое наблюдавшихся значений величины Y , соответствующих $X=x$. Функция изменения условного среднего $\bar{y}_x = f(x)$ от независимой переменной X называется уравнением регрессии. Уравнения регрессии строятся для зависимых переменных. Эти переменные входят в левую часть уравнения. Независимые переменные входят в правую часть уравнения и позволяют предсказывать зависимую переменную.

Предсказанные значения зависимой переменной – значения \bar{y}_x , вычисленные по уравнению регрессии с оцененными коэффициентами регрессии. Остатки – разности между наблюдаемыми и предсказанными значениями зависимой переменной:

$$\varepsilon_i = y_i - \bar{y}_x(x_i), \quad (2.4)$$

Сумма квадратов остатков – сумма вида:

$$S_\varepsilon = \sum_{i=1}^n [y_i - \bar{y}_x(x_i)]^2 \quad (2.5)$$

Сумма квадратов зависимой переменной, скорректированная на среднее \bar{y}

$$S = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.6)$$

Сумма квадратов предсказанной зависимой переменной, скорректированная на среднее \bar{y}

$$S_x = \sum_{i=1}^n [\bar{y}_x(x_i) - \bar{y}]^2 \quad (2.7)$$

Известно, что для суммы квадратов указанных величин, выполняется равенство:

$$S = S_x + S_\varepsilon. \quad (2.8)$$

Коэффициент детерминации и скорректированный коэффициент детерминации

$$RI = \frac{S_x}{S}, \quad RI_S = 1 - (1 - RI) \cdot \frac{n}{n - p},$$

где: n – число наблюдений, p – число параметров модели (число независимых переменных плюс 1, так как обычно в модель включается свободный член).

Наибольшее применение получили уравнения регрессии, отражающие взаимосвязь одной зависимой переменной с одной (парная регрессия) или несколькими (множественная регрессия) независимыми переменными.

Чаще всего используют следующие парные и множественные зависимости:

1) Парная и множественная линейная регрессия:

$$Y = a_0 + a_1 \cdot X, \quad Y = a_0 + \sum_{n=1}^N a_n \cdot X_n \quad (2.9)$$

2) Парная и множественная параболическая регрессия:

$$Y = a_0 + a_1 \cdot X + a_2 \cdot X^2, \quad Y = a_0 + \sum_{n=1}^N (a_n \cdot X_n + b_n \cdot X_n^2) \quad (2.10)$$

3) Парная и множественная гиперболическая регрессия:

$$Y = a_0 + a_1/X, \quad Y = a_0 + \sum_{n=1}^N a_n/X_n \quad (2.11)$$

4) Парная и множественная степенная регрессия:

$$Y = a_0 \cdot X^k, \quad Y = a_0 \cdot X_1^{k1} \cdot X_2^{k2} \dots \cdot X_n^{kn} \quad (2.12)$$

5) Парная и множественная показательная регрессия:

$$Y = a_0 \cdot a_1^X, \quad Y = a_0 \cdot a_1^{X1} \cdot a_2^{X2} \dots \cdot a_n^{Xn}, \quad Y = a_0 + \sum_{n=1}^n b_n \cdot a_n^{Xn} \quad (2.13)$$

Обычно пытаются использовать линейные зависимости или зависимости, которые приводят к пути преобразования переменных. Параметры уравнения регрессии выбираются методом наименьших квадратов. Он предоставляет минимальную сумму квадратов отклонений фактических значений Y от рассчитанных с использованием уравнения регрессии для заданных значений независимых переменных.

Регрессионный анализ данных предполагает, что выбирается наиболее оптимальный вид функции регрессии $f(x_1, x_2, x_3, \dots)$ из набора (2.9) – (2.13), оцениваются коэффициенты функции регрессии a_0, a_1, a_2, \dots и строятся для них доверительные интервалы, проверяется гипотеза о значимости регрессии, оценивается степень адекватности модели и т.д.

Обычно выбор уравнения регрессии осуществляется поэтапно. На первом этапе выбираются зависимая переменная и одна из самых мощных независимых переменных, полученные по результатам корреляционного анализа. Далее строим пару зависимостей, определяем коэффициент корреляции и его значимость. На втором шаге добавляют следующую весовую переменную и строят уравнение регрессии для зависимой переменной Y на двух выбранных независимых переменных. Коэффициент множественной корреляции определяется, и регрессия оценивается. Далее, если необходимо, добавляется следующая переменная и т. д. Возможный путь возврата связан с постепенным устранением малозначных переменных. На каждом этапе выполняется графический анализ данных, исключаются некоторые ненормальные наблюдения и оценивается значимость регрессии. Оценка адекватности модели осуществляется путем применения различных процедур для анализа распределения остатков.

Увеличение размерности уравнений регрессии увеличивает значение коэффициента детерминации. Однако увеличивать размерность (более 2 – 3-х переменных в модели) путем добавления новых независимых переменных

имеет смысл, когда наблюдается явное улучшение показателей регрессии: увеличение коэффициента детерминации RI и уменьшение суммы квадратов остатков S_ε . Следует придерживаться общего правила, что не следует гнаться за чрезмерной сложностью модели.

Для оценки значимости уравнения регрессии в целом применяют F -критерий:

$$F_{рас} = \frac{RI}{1 - RI} \cdot \frac{n - p}{p - 1}.$$

F -критерий используется для проверки гипотезы о значимости регрессии. Она утверждает, что между зависимой переменной и независимыми переменными нет линейной связи, то есть что коэффициенты регрессии равны нулю, против альтернативы, что они не равны нулю. Для проверки гипотезы расчетное значение F -критерия сравнивается с табличным значением F -критерия при уровне значимости α и $k_1 = p - 1, k_2 = n - p$ степенях свободы. Если $F_{рас} > F_\alpha(k_1, k_2)$, то уравнение регрессии можно признать статистически значимым, т.е. гипотеза о значимости регрессии подтверждается.

Оценка значимости независимых переменных осуществляется на основе t -критерия

$$t_{i,рас} = \frac{a_i}{\sigma_i},$$

где: a_i – численное значение i -того коэффициента уравнения множественной регрессии;

σ_i – среднеквадратичное отклонение параметра a_i как случайной величины относительно среднего уровня.

Расчетное значение t -критерия сравнивают по абсолютной величине с табличным значением t -критерия при заданном уровне значимости α и $k = n - p$ степенях свободы. Если $t_{i,рас} > t_\alpha(k)$, то параметр a_i считается

значимым и соответствующая независимая переменная отбирается в уравнение множественной регрессии.

В случае если данные подчинены нелинейной связи, то используют преобразованные переменные: $\ln(y)$ и $\ln(x)$ для уравнений вида (2.12), y и $1/x$ для уравнений вида (2.11), $\ln(y)$ и x для уравнений вида (2.13), y и x, x^2 для уравнений вида (2.10) и т.д. При таком преобразовании переменных задача сводится к определению линейного уравнения регрессии вида (2.9).

В практике комплексной оценки используются методы индексной оценки состояния российских регионов в виде интегрального индекса качества жизни населения, который имеет следующую форму [24]:

$$ИИКЖ = \frac{1}{8} \sum_{i=1}^8 I_k \quad (2.14)$$

где I_k – значение частного индекса качества жизни населения.

Расчёт производится по формуле:

$$I_k = \frac{x - x_{\min}}{x - x_{\max}} \quad (2.15)$$

где I_k – значение частного индекса качества жизни населения, x – оцениваемый показатель, x_{\max} и x_{\min} – максимальные и минимальные значения показателей среди регионов страны. Данная методика позволяет произвести сравнительный анализ регионов и проследить динамику [24].

Существует также метод сравнения регионов России с точки зрения человеческого развития, отражающий уровень жизни, грамотность, образование и долголетие. Эти показатели являются основной характеристикой человеческого потенциала и присущи каждому региону. При расчёте индекса человеческого развития, так называемого ИЧР, учитываются четыре показателя: ожидаемая продолжительность жизни, уровень грамотности населения страны и ожидаемая продолжительность

обучения; уровень жизни оценивается в единицах валового национального дохода на душу населения по паритету покупательной способности в долларах США.

Для расчётов используют формулу (2.15) для индикаторов и определяют частные индексы:

1. индекс ожидаемой продолжительности жизни:

$$I_1 = \frac{x_1 - 20}{83,2 - 20} \quad (2.16)$$

2. индекс образования:

$$I_2 = \frac{\sqrt{i_2 i_3 - 0}}{0,951 - 0}, \quad (2.17)$$

где индекс средней продолжительности обучения:

$$i_2 = \frac{x_2 - 0}{13,2 - 20}, \quad (2.18)$$

а индекс ожидаемой продолжительности обучения:

$$i_3 = \frac{x_3 - 0}{20,6 - 0}, \quad (2.19)$$

3. индекс дохода:

$$I_4 = \frac{\ln x_4 - \ln 163}{\ln 108211 - \ln 163} \quad (2.20)$$

ИРЧП является средней геометрической этих трёх индексов:

$$HDI = \sqrt{I_1 I_2 I_4} \quad (2.21)$$

Индекс развития человеческого потенциала на региональном уровне содержит важные показатели, которые характеризуют уровень развития человека, от которого зависит развитие регионов и всей страны в целом.

3 ПРОЕКТИРОВАНИЕ WEB-ПРИЛОЖЕНИЯ ДЛЯ АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ

В ходе изучения различных веб-приложений для анализа статистических данных нами были установлены требования и необходимые функциональные возможности, подлежащие реализации в виде веб-приложения для анализа многомерных данных. Структурная декомпозиция будет рассмотрена ниже.

Так же, в ходе проектирования был выбран набор средств и технологий реализации данного приложения, позволяющий разработать продукт в наибольшей степени удовлетворяющий сформированным нами требованиям.

3.1 Выбор технологий и средств реализации

Для ведения разработки в рамках локальной версии приложения нами используется локальный сервер Open Server.

Open Server — это портативный локальный WAMP/WNMP сервер, являющийся многофункциональной управляющей программой и предоставляющий широкий спектр подключаемых компонентов.

Open Server — это полностью портативный сервер. Он не требует дополнительных системных сервисов, не создаёт мусор в реестре и папке system32. Система может быть перемещена на съёмном носителе (желательно на скоростном флеш-накопителе), и запущена на любой рабочей/домашней машине.

В случае отсутствия на компьютере нужных системных компонентов Open Server установит их сам, достаточно выбрать в меню (Инструменты — Первый запуск) если сервер запускается на компьютере впервые.

Нами выбран Open Server, так как это наиболее удобный профессиональный инструмент, созданный специально для веб-разработчика.

Вести разработку web-приложения мы будем с использованием языка программирования PHP. Данный язык наиболее часто применяется в web-разработке, так как является наиболее гибким и адаптируемым в данной области применения.

В качестве СУБД используем MySQL, также наиболее распространённое средство для ведения web-разработки.

PhpStorm это интеллектуальная среда разработки для PHP, HTML и JavaScript с возможностью анализа кода на лету, предотвращения ошибок в коде и встроенными автоматизированными средствами рефакторинга для PHP и JavaScript. PhpStorm поддерживает автодополнения кода, с учётом спецификации всех современных и традиционных проектов PHP, включая генераторы, сопрограммы, пространства имён, замыкания, типаж и синтаксис коротких массивов. В IDE встроен полноценный SQL-редактор, имеющий возможность редактирования полученных результатов запросов.

MySQL Workbench – инструмент для визуального проектирования баз данных, интегрирующий проектирование, моделирование, создание и эксплуатацию БД в единое бесшовное окружение для системы баз данных MySQL.

MySQL Workbench даёт нам следующие возможности:

- доступный визуальный интерфейс для моделирования базы данных;
- предоставляет удобные и функциональные средства построения связей между таблицами, включая связи вида «многие ко многим» с созданием таблицы связей;
- Reverse Engineering – восстановление структуры таблиц из уже существующей на сервере БД (связи восстанавливаются в InnoDB, при использовании MyISAM — связи необходимо устанавливать вручную);
- удобный механизм создания SQL запросов, позволяющий отправлять их серверу и увидеть ответ в виде таблицы;

— осуществлять редактирование данных в таблице в визуальном режиме.

В качестве средства автоматизации программирования мы будем использовать API для разработки web-приложений Yii.

Это объектно-ориентированный компонентный фреймворк, разработанный на PHP и реализующий парадигму MVC. Yii - это универсальная среда веб-программирования, использующая PHP. Он особенно подходит для разработки крупномасштабных приложений, таких как порталы, форумы, системы управления контентом (CMS), проекты электронной коммерции, веб-сервисы RESTful и так далее.

Его основные возможности:

- парадигма MVC (Модель-представление-контроллер),
- интерфейсы БД DAO и ActiveRecord (PDO),
- кэширование страниц и отдельных фрагментов,
- аутентификация и авторизация (rbac и acl),
- инструменты ввода и валидации форм,
- генерация базового php-кода для CRUD-операций (скаффолдинг),
- отслеживание и обработка ошибок,
- миграции базы данных,
- поддержка ajax и интеграция с jquery,
- поддержка интернационализации,
- поддержка тем оформления для их лёгкой смены,
- автоматическое тестирование,
- возможность подключения сторонних библиотек,
- поддержка REST.

Как и большинство PHP-фреймворков, Yii реализует архитектурный шаблон MVC (Model-View-Controller) и способствует организации кода на основе этого шаблона.

Yii придерживается философии, что код должен быть написан простым, но элегантным способом. Yii никогда не будет пытаться перепроектировать вещи главным образом с целью строго следовать некоторому шаблону дизайна.

Yii - это полнофункциональная платформа, предоставляющая множество проверенных и готовых к использованию функций: строители запросов и ActiveRecord для реляционных баз данных и баз данных NoSQL; Поддержка разработки RESTful API; поддержка многоуровневого кэширования; и больше.

Yii чрезвычайно расширяемый. Можно настроить или заменить почти каждый фрагмент кода ядра. Также можно воспользоваться преимуществами архитектуры расширений Yii для использования или разработки распространяемых расширений.

Высокая производительность всегда является основной целью Yii.

Yii - это не персональное ПО, его поддерживают сильная команда разработчиков ядра, а также большое сообщество профессионалов, постоянно вносящих свой вклад в развитие Yii. Команда разработчиков Yii внимательно следит за последними тенденциями веб-разработки, а также за лучшими практиками и функциями, которые можно найти в других средах и проектах. Наиболее важные передовые практики и функции, встречающиеся в других местах, регулярно включаются в базовую структуру и представляются через простые и элегантные интерфейсы.

3.2 Определение функций и задач модулей системы

В соответствии с поставленной задачей нам необходима система (web-приложение) осуществляющая анализ многомерных данных, характеризующая города и регионы России. Организационная структура такой системы соответствует диаграмме на рисунке 3.1. Как видно из

диаграммы система рассматривает алгоритмы методов математического анализа как руководящие инструкции.



Рис 3.1 – Общая организационная диаграмма функционирования web-системы анализа многомерных данных

На рисунке 3.2 приведена диаграмма декомпозиции процесса анализа многомерных данных. Как видно из рисунка в функции системы входит приём входных данных, их хранение, обработка и вывод в доступной для пользователя форме, что выделено в три отдельных независимых процесса.

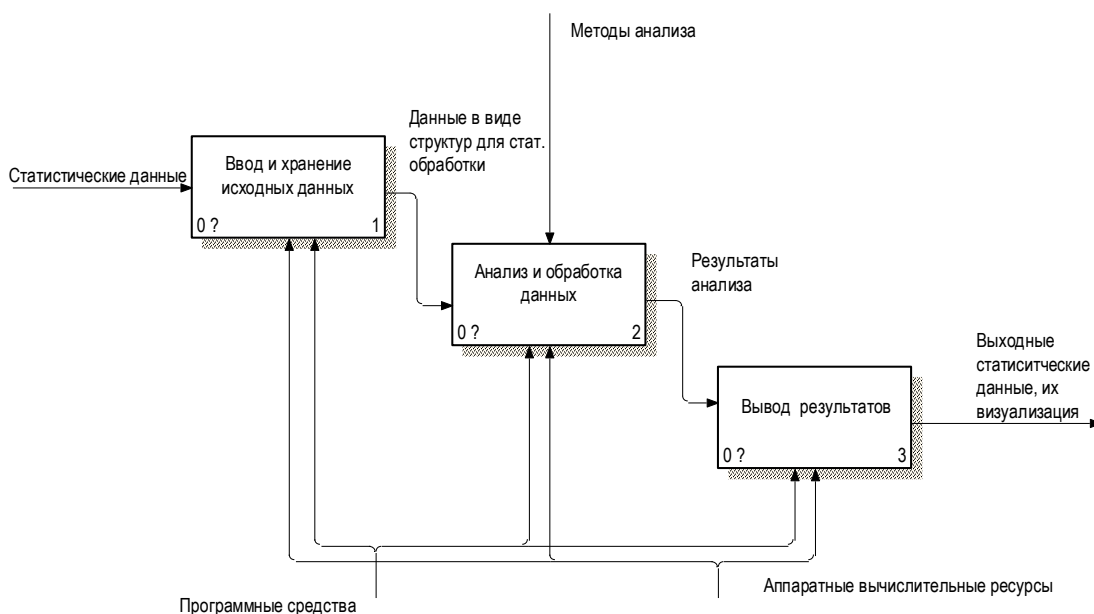


Рис. 3.2 – Структурная декомпозиция основного процесса функционирования системы

Поступление вводных данных в систему может происходить различными путями. Оптимально, когда данные импортируются при помощи файла. Но иногда необходим ручной ввод. Логическая схема процесса ввода данных изображена на рисунке 3.3

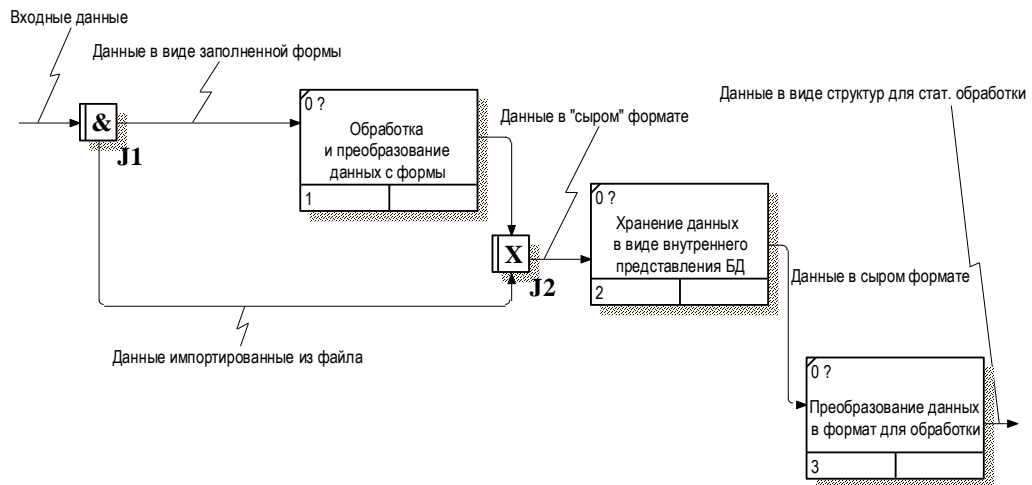


Рис. 3.3 – Схема процесса ввода и хранения исходных данных

Исходя из наших исследований в предыдущих главах данной работы, мы опираемся на два основных математических метода анализа данных. На рисунке 3.4 приведена соответствующая логическая модель обработки данных.

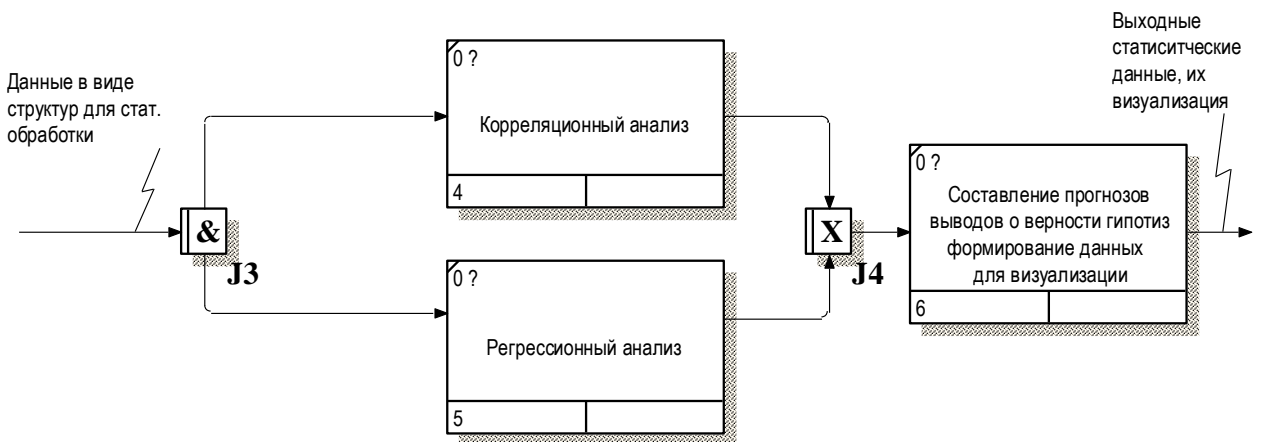


Рис. 3.4 – Процесс анализа данных

Как видно из приведённой схемы отдельным блоком осуществляется анализ и прогнозирование на основе проведённого анализа.

На рисунке 3.4 приведена диаграмма процесса формирования выходных данных, которые будут получены пользователем.

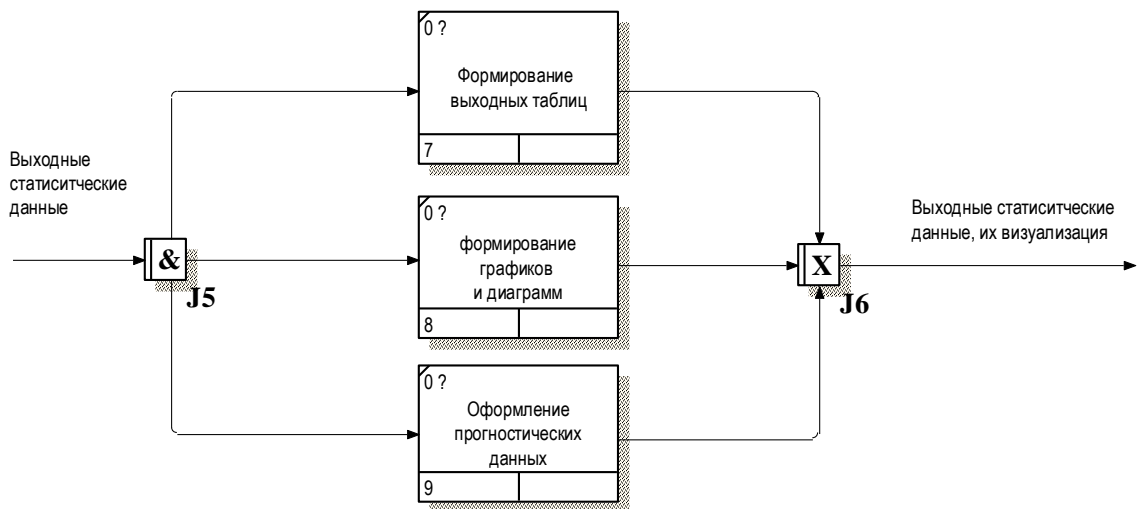


Рис. 3.4 – Процесс формирования выходных данных

3.3 Разработка структуры приложения и методы реализации

В рамках реализации web-приложения в проект закладывается структура приложения приведённая на рисунке 3.5.

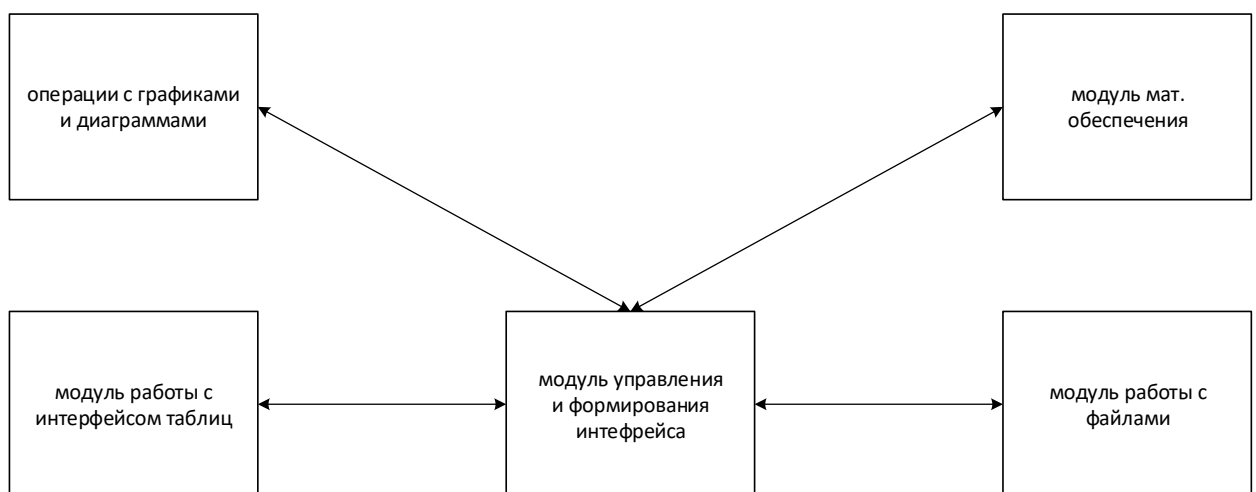


Рис. 3.5 – Структурная схема приложения

Как видно из схемы, приложение разбивается на пять основных функциональных блоков:

- модуль управления программой и формирования пользовательского интерфейса.
- модуль работы с интерфейсом таблиц
- модуль операций с графиками и диаграммами
- модуль работы с файлами
- модуль математического обеспечения.

Для хранения и оптимального доступа к данным предлагается максимально декомпозировать данные. На рисунке 3.6 приведена соответствующая реляционная модель данных

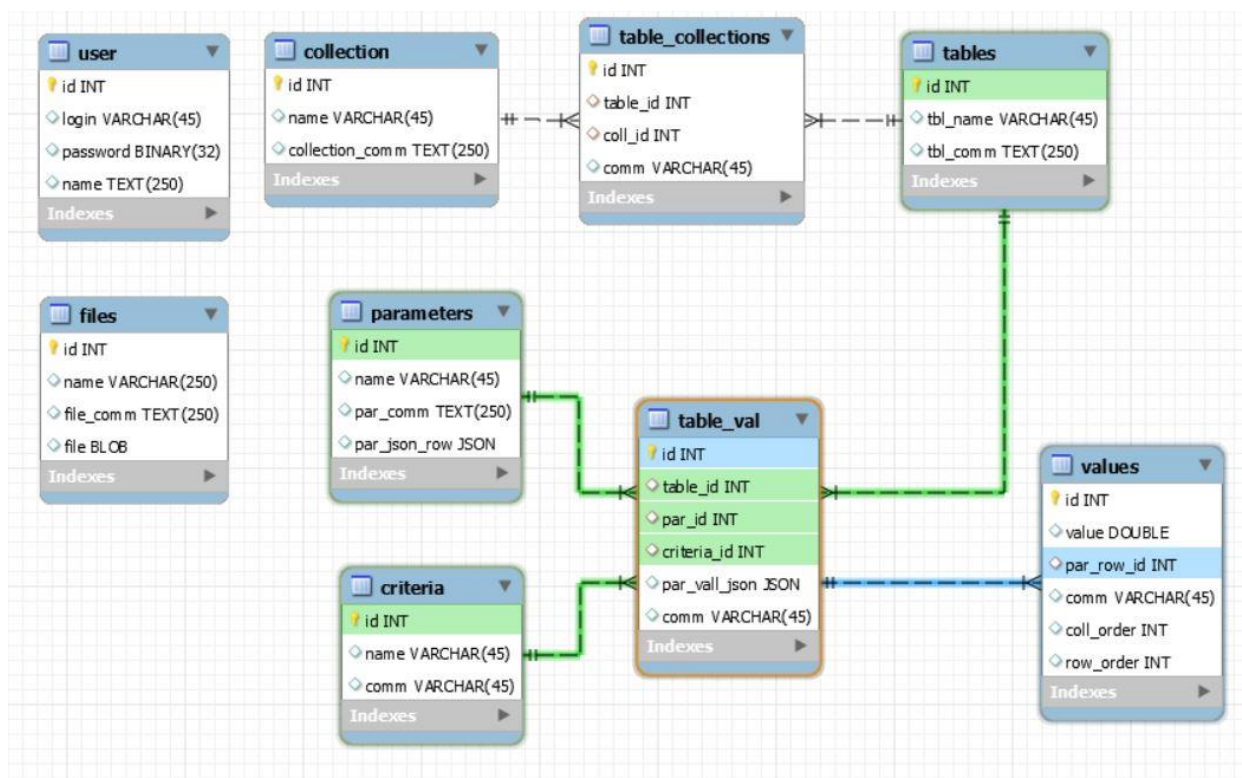


Рис. 3.6 – Структура данных системы

Для создания различных наборов данных для анализа применяется

сущность «коллекция». Коллекция представляет собой определённый набор данных для обработки. Это могут быть данные из нескольких таблиц или единичного вектора.

Параметр – это набор значений определённого вектора данных. Например – значения ежегодных наблюдений по годам.

Критерий – это определённые точки фиксации значения параметра, например, годы. Критерии соответствуют столбцам таблицы, параметры – строкам, а сущность «значение» ячейкам таблицы.

3.4 Реализация приложения.

На рисунке 3.7 изображён разработанный интерфейс меню приложения, оно содержит пункты по которым доступны данные статистики.

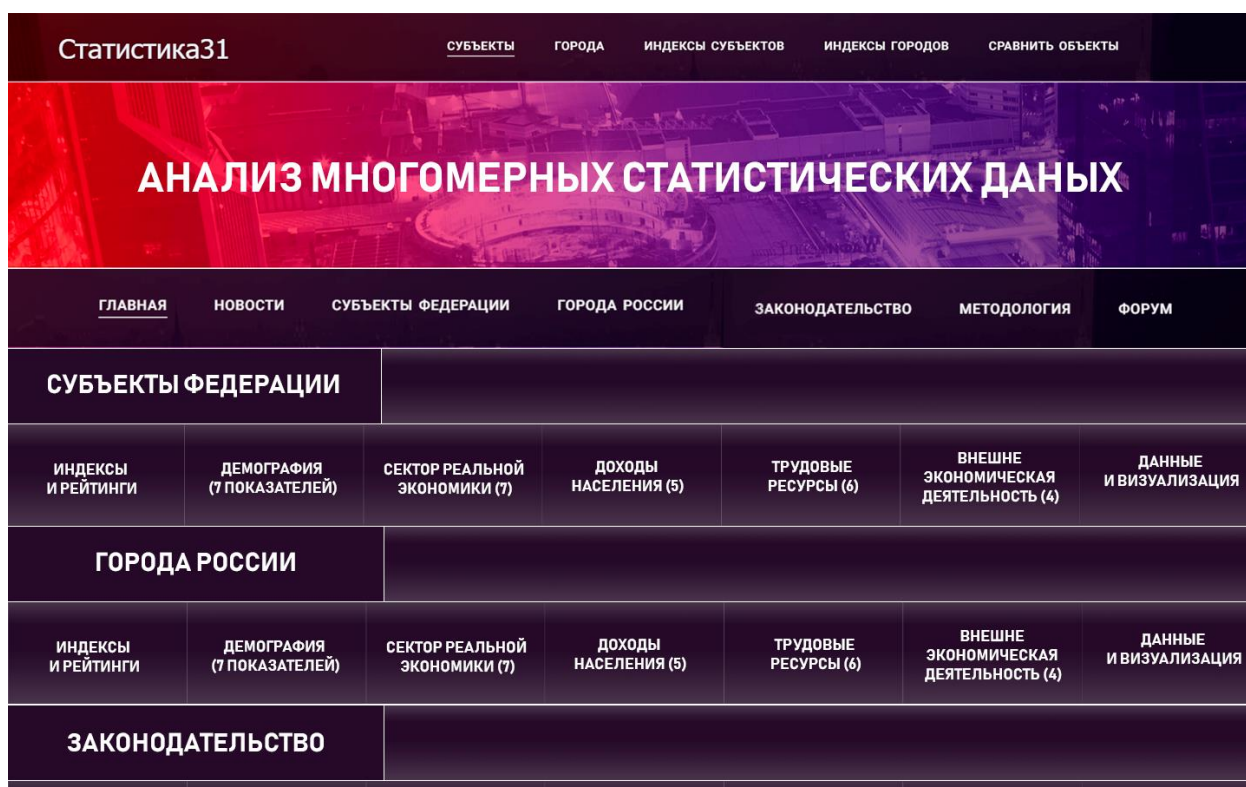


Рис. 3.7 – Интерфейс меню приложения

СТРУКТУРА ПРОИЗВОДСТВА ПО ОТРАСЛЯМ

УРОВЕНЬ НАЦИОНАЛЬНОГО ДОХОДА (НД)

УРОВЕНЬ НАЦИОНАЛЬНОГО БОГАТСТВА (НБ)

УРОВЕНЬ ЧИСТОГО ЭКОНОМИЧЕСКОГО БЛАГОСОСТОЯНИЯ (ЧНБ)

ВВП (ВАЛОВЫЙ ВНУТРЕННИЙ ПРОДУКТ)

ВНП (ВАЛОВЫЙ НАЦИОНАЛЬНЫЙ ПРОДУКТ)

ДОБАВИТЬ В АНАЛИЗ

Рис. 3.8 – Выбор данных для анализа

КОЭФФИЦИЕНТ СУММАРНОЙ РОЖДАЕМОСТИ

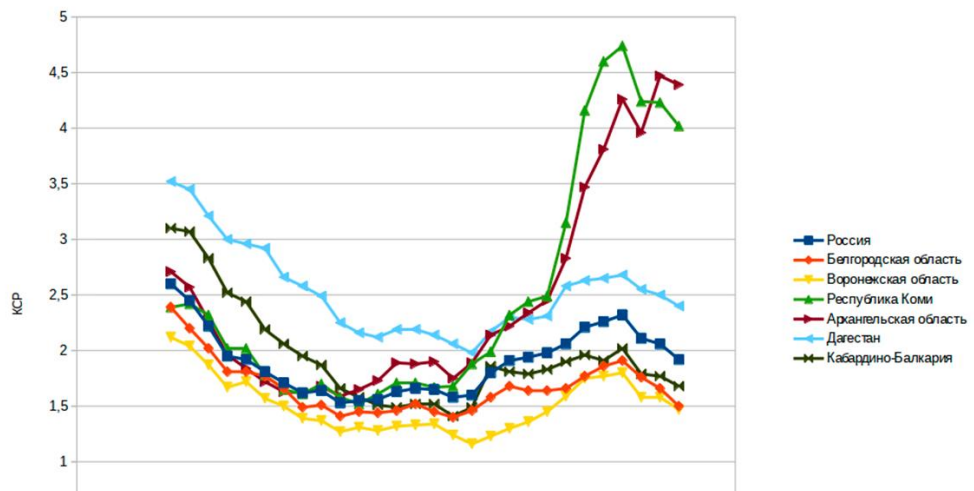


Рис. 3.9 – Результат анализа.

На рисунке 3.8 изображён процесс выбора набора данных для анализа. Рисунок 3.9, в свою очередь приводит график, полученный в результате анализа.

ЗАКЛЮЧЕНИЕ

Были изучены существующие официальные ресурсы различных организаций, деятельность которых связана со сбором, накоплением и обработкой статистической информации о состоянии и развитии городов и регионов по различным показателям.

Рассматриваются существующие сайты, предоставляющие данные по различным показателям разных городов и регионов. Рассматриваются базы данных, описывающие состояние и развитие городов и регионов в различных аспектах их деятельности. Для дальнейших исследований была выбрана база данных Федеральной службы государственной статистики. Были собраны статистические данные о состоянии и развитии городов России.

Анализируются существующие меры сходства объектов, в качестве методов рассматриваются евклидово расстояние, манхэттенское расстояние, расстояние Чебышева, расстояние экспоненты и методы расположения центра тяжести.

Выполнен кластерный анализ состояния регионов России по пять показателям. Установлено значение пяти кластеров, в которые входят 79 регионов. Обоснован выбор подходов для анализа статистических данных в виде пяти математических зависимостей и нескольких комплексных индексов.

На основе полученных данных был разработан и протестирован веб-ресурс для анализа многомерных статических данных с использованием представленных математических моделей. С этой целью выбраны технологии и средства реализации: локальный сервер Open Server, язык программирования PHP и СУБД MySQL.

Определены функции и задачи модулей системы, представлены схемы функционирования, ввода, хранения и анализа данных, процессов

формирования выходных данных. Разработана структурная схема приложения из пяти модулей.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Федеральная служба государственной статистики России
<http://www.gks.ru/>

2. Стратегическое планирование в Российской Федерации: состояние методического обеспечения. Аналитический доклад (по результатам мониторинга реализации Федерального закона от 28.06.2014, №172-ФЗ «О стратегическом планировании в Российской Федерации»): препринт WP8/2016/02 / А.В. Клименко, В.А. Королев, Д.Ю. Двинских, Н.А. Рычкова, И.Ю. Сластихина. – М.: Вс. шк. экономики, 2016. – 60 с.

3. Методические рекомендации по разработке, корректировке, мониторингу среднесрочного прогноза социально-экономического развития Российской Федерации». Утв. приказом Минэкономразвития России [от 30.06.2016, №423] // [Электронный ресурс]. – Режим доступа:<http://docs.cntd.ru/document/420383543> (28.05.2018).

4. Методические рекомендации по подготовке стратегий развития отраслей экономики [Электронный ресурс]. – Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_256544/3286efd0adaba9d26d5320d40f074bff8ceacfb6/ (02.06.2018).

5. Методические рекомендации по оптимизации стратегического планирования на муниципальном уровне. М.: Институт экономики города, 2015. – 32 с.

6. Методические рекомендации по согласованной подготовке и реализации документов планирования развития муниципальных образований / Л.Ю. Падиляя[и др.]. – М.: Институт экономики города, 2010. – 112 с.

7. Методические рекомендации к разработке показателей прогнозов социально-экономического развития субъектов Российской Федерации [Электронный ресурс]. – Режим доступа: <http://samzan.ru/63030> (01.06.2018).

8. Методические рекомендации по заполнению формы (форма 2П) и к разработке показателей прогнозов социально-экономического развития субъектов Российской Федерации [Электронный ресурс]. – Режим доступа: http://volgafin.volgograd.ru/upload/iblock/0ba/6_metod_rekomendatzii.pdf (02.06.2018).

9. Методические рекомендации по разработке и корректировке стратегии социально-экономического развития субъекта Российской Федерации и плана мероприятий по ее реализации. Утв. приказом Минэкономразвития России. 2017 г.

10. Перечень показателей мониторинга процессов в реальном секторе экономики, финансово-банковской и социальной сферах субъектов РФ: распоряж. Правительства РФ от 15.06.2009, №806-р (ред. от 13.12.2017). [Электронный ресурс]. – Режим доступа: http://www.consultant.ru/document/cons_doc_LAW_2077_36/ (25.05.2018).

11. Звягинцева А.В. Вероятностные методы комплексной оценки природно-антропогенных систем / под ред. проф. Г.В. Аверина, 2016. – М.: Спектр. – 257 с.

12. Сошникова Л.А. Многомерный статистический анализ в экономике/ Сошникова Л.А., Тимашевич В.Н., Уебе Г., Шефер М.: уч. пос. для вузов/ под ред. Проф. В.Н. Тимашевича. – М.: ЮНИТИ-ДАНА, 1999. – 598 с.

13. Дейвисон М. Многомерное шкалирование: Методы наглядного представления данных / Пер. с англ. – М.: Финансы и статистика, 1988. – 254 с.

14. Прогнозирование будущего. Новая парадигма / под ред. Г.Г. Фетисова, В.М. Бондаренко. – М.: Экономика, 2008. – 283 с.

15. Econophysics and sociophysics: trends and perspectives /B.K. Chakrabarti, A. Chakraborti, A. Chatterie (eds.). Berlin: Wiley-VCH, 2006, 622 p.
16. Гирина А.Н. Методика оценки социально-экономического развития региона // Вестник ОГУ. 2013, №8(157). – С. 82–87.
17. Россошанский А.И. Методика индексной оценки качества жизни населения российских регионов // Проблемы развития территории. 2016, № 4. – С. 124–137.
18. Аверин Г.В., Звягинцева А.В. О справедливости принципа соответственных состояний для систем различной природы // Научные ведомости Белгородского государственного университета. Сер. Экономика. Информатика, №16(265), вып. 43. 2017. – С. 104–112.
19. Печерских И.А., Семенов А.Г. Математические модели в экономике: уч. пос. для студ. и вузов. – Кемерово, 2011. – 191 с.
20. Еремеев В.Н. Моделирование эколого-экономических систем / В.Н. Еремеев, Е.М. Игумнова, И.Е. Тимченко. Севастополь: ЭКОСИ-гидрофизика, 2004. – 322 с.
21. Mathematical modeling of collective behavior in socio-economic and life sciences / G. Naldi, L. Pareschi, G. Toskani (eds.). Berlin, Springer, 2010, 438 p.
22. Цыгичко В.Н. Прогнозирование социально-экономических процессов. 2017. – 240 с.
23. Гмурман В.Е. Теория вероятности и математическая статистика. / Уч. пос. для вузов. М.: Вс. шк., 1998. – 479 с.
24. Россошанский А.И. Методика индексной оценки качества жизни населения российских регионов // Проблемы развития территории. 2016, № 4. – С. 124–137.

ПРИЛОЖЕНИЕ А

Фрагмент исходного кода класса модели документа

```
<?php
```

```
class ModelDocumentDocument extends Model {

    private $step = 0;
    private $recursion = array();
    private $step_view = 0;

    public function addDocument($doctype_uid, $author_uid = 0, $draft = 3) {
        if (!$author_uid) {
            $author_uid = $this->customer->getStructureId();
        }
        $this->load->model('account/customer');
        $department_uid = $this->model_account_customer->getParentStructure($author_uid);
        $query = $this->db->query("SELECT document_uid FROM " . DB_PREFIX . "document WHERE draft = 3 AND author_uid = '" . $this->db->escape($author_uid) . "' AND doctype_uid = '" . $this->db->escape($doctype_uid) . "'");
        if ($query->num_rows) {
            $document_uid = $query->row['document_uid'];
        } else {
            $query = $this->db->query("SELECT UUID() AS uid");
            $document_uid = $query->row['uid'];
        }
    }
}
```

```

$this->db->query("INSERT INTO " . DB_PREFIX . "document SET "
    . "document_uid = " . $document_uid . ", "
    . "doctype_uid = " . $this->db->escape($doctype_uid) . ", "
    . "author_uid = " . $this->db->escape($author_uid) . ", "
    . "department_uid = " . $this->db->escape($department_uid) . ", "
    . "route_uid = " . $this->getFirstRoute($doctype_uid) . ", "
    . "draft = " . (int) $draft . ", "
    . "date_added=NOW()"
);

```

```

$this->db->query("INSERT INTO " . DB_PREFIX . "document_access
SET subject_uid=" . $this->db->escape($author_uid) . ", document_uid = " .
$this->db->escape($document_uid) . ", doctype_uid = " . $this->db-
>escape($doctype_uid) . " ");

```

```

}
return $document_uid;
}

```

```

* @param type $document_uid
*/

```

```

public function editDocument($document_uid) {

```

```

    $this->db->query("UPDATE " . DB_PREFIX . "document SET date_added =
(CASE WHEN draft=3 THEN NOW() ELSE date_added END), draft=0,
draft_params=", date_edited=NOW() WHERE document_uid=" . $this->db-
>escape($document_uid) . " ");

```

```

}

```

```

/**

```

```

* @param type $document_uid

```



```

*/
public function saveDraftDocument($document_uid, $data) {
    $this->db->query("UPDATE " . DB_PREFIX . "document SET draft=(CASE
WHEN draft=3 THEN 3 ELSE 1 END), draft_params=" . $this->db-
>escape(serialize($data)) . " WHERE document_uid=" . $this->db-
>escape($document_uid) . " ");
}

public function removeDraftDocument($document_uid) {
    $this->db->query("UPDATE " . DB_PREFIX . "document SET draft=(CASE
WHEN draft=3 THEN 3 ELSE 0 END), draft_params=" WHERE
document_uid=" . $this->db->escape($document_uid) . " ");
    $document_info = $this->getDocument($document_uid);
    if ($document_info['draft'] == 3) {
        $this->removeDocument($document_uid);
    }
}

/**
 * Возвращает информацию о документах
 * @param type $document_uid
 * @param type $check_access – проверяется доступ к документам для
текущего пользователя
 * @return type
 */
public function getDocument($document_uid, $check_access = true) {
    if ($check_access) {

```

```

$query = $this->db->query("SELECT DISTINCT d.doctype_uid,
d.author_uid, d.department_uid, d.route_uid, d.draft, d.draft_params, d.date_added,
d.date_edited, "
    . "dt.field_log_uid FROM " . DB_PREFIX . "document d "
    . "LEFT JOIN doctype dt ON (dt.doctype_uid = d.doctype_uid) "
    . "WHERE d.document_uid = '" . $this->db->escape($document_uid) .
    "' "
    . "AND ("
    . " ((SELECT doctype_uid FROM " . DB_PREFIX . "doc-
ument_access WHERE document_uid='" . $this->db->escape($document_uid) . "'
AND subject_uid IN (" . implode(",", $this->customer->getStructureIds()) . ")
LIMIT 0,1) IS NOT NULL) "
    . " OR "
    . " ((SELECT doctype_uid FROM " . DB_PREFIX . "doc-type_access
WHERE doctype_uid=d.doctype_uid AND (object_uid=d.author_uid OR
object_uid=d.department_uid) AND subject_uid IN (" . implode(",", $this-
>customer->getStructureIds()) . ") LIMIT 0,1) IS NOT NULL) "
    . ") "
    );
} else {
$query = $this->db->query("SELECT DISTINCT d.doctype_uid,
d.author_uid, d.department_uid, d.route_uid, d.draft, d.draft_params, d.date_added,
d.date_edited, dt.field_log_uid FROM " . DB_PREFIX . "document d "
    . "LEFT JOIN doctype dt ON (dt.doctype_uid = d.doctype_uid) "
    . "WHERE d.document_uid = '" . $this->db->escape($document_uid) .
    "'");
}
return $query->row;
}

```

```

/**
 *
 */

/**
 * Метод возвращает идентификаторы документов на основании критерием
в $data
 * @param type $data:
 * 'author_uids' – массив с идентификаторами авторов документов
 * 'doctype_uids' – массив с идентификаторами типов документов
 * 'document_uids' – выборка из массива с document_uid
 * 'filter_names' – условия по значению полей: filed_id =>
$condition_value = array($condition,$value) (value может быть заменен на
display, чтобы выборка шла по display
 */

public function getDocumentIds($data) {
    $this->load->model('doctype/doctype');
    $joins = array();
    $where = "";
    $sql = "SELECT ";
    if (!empty($data['function'])) {
        $sql .= $this->db->escape($data['function']) . "(";
        if (!empty($data['function_join'])) {
            $field_info = $this->model_doctype_doctype-
>getField($data['function_join']);
            if ($field_info) {
                $sql .= "fv" . $this->db->escape(str_replace("-", "", $da-
ta['function_join'])) . ".value";

```

```

        $joins[] = "LEFT JOIN " . DB_PREFIX . "field_value_" . $this->db-
>escape($field_info['type']) . " fv" . $this->db->escape(str_replace("-", "",
$data['function_join'])) . " ON (fv" . $this->db->escape(str_replace("-", "",
$data['function_join'])) . ".document_uid=d.document_uid AND fv" . $this->db-
>escape(str_replace("-", "", $data['function_join'])) . ".field_uid = " . $this->db-
>escape($data['function_join']) . ") ";
    }
} else {
    $sql .= "d.document_uid";
}
$sql .= ") AS result ";
} else {
    $sql .= "d.document_uid ";
}
$sql .= "FROM " . DB_PREFIX . "document d ";
if (!empty($data['filter_names'])) {
    foreach ($data['filter_names'] as $field_uid => $condition_value) {
        if (count($condition_value)) {
            $field_info = $this->model_doctype_doctype->getField($field_uid);
            if (!$field_info) {
                continue;
            }
            $joins[] = "LEFT JOIN " . DB_PREFIX . "field_value_" . $this->db-
>escape($field_info['type']) . " fv" . $this->db->escape(str_replace("-", "",
$field_uid)) . " ON (fv" . $this->db->escape(str_replace("-", "", $field_uid)) .
".document_uid=d.document_uid AND fv" . $this->db->escape(str_replace("-", "",
$field_uid)) . ".field_uid = " . $this->db->escape($field_uid) . ") ";
            foreach ($condition_value as $condition) {
                if (isset($condition['value'])) {

```

```

$stable_field = "value";
$value = $condition['value'];
} else {
    $stable_field = "display_value";
    $value = $condition['display'];
}
//проверяю знаки сравнения
$add_cond = "";
$link_add_cond = "OR";
switch ($condition['comparison']) {
    case '=':
        $comparison = '=';
        if (!$value) {
            $add_cond = "IS NULL";
        }
        break;
    case '>':
        $comparison = '>';
        break;
    case '<':
        $comparison = '<';
        break;
    case 'equal':
        $comparison = '=';
        if (!$value) {
            $add_cond = "IS NULL";
        }
        break;
    case 'notequal':

```

```

$comparison = '<>';
if ($value) {
    $add_cond = "IS NULL";
} else {
    $add_cond = "IS NOT NULL";
    $link_add_cond = "AND";
}
break;
case 'more':
    $comparison = '>';
    break;
case 'moreequal':
    $comparison = '>=';
    break;
case 'less':
    $comparison = '<';
    $add_cond = "IS NULL";
    break;
case 'lessequal':
    $comparison = '<=';
    break;
case 'contains':
    $comparison = 'LIKE';
    $value = "%" . $value . "%";
    break;
case 'notcontains':
    $comparison = 'not LIKE';
    if ($value) {
        $add_cond = "IS NULL";
    }

```

```

    } else {
        $add_cond = "IS NOT NULL";
        $link_add_cond = "AND";
    }
    $value = "%" . $value . "%';

    break;
default:
    $comparison = '=';
    break;
}

$wh = "(fv" . $this->db->escape(str_replace("-", "", $field_uid)) .
"." . $table_field . " " . $comparison . " " . $this->db->escape($value) . "" .
($add_cond ? " " . $link_add_cond . " fv" . $this->db->escape(str_replace("-", "", $field_uid)) . "." . $table_field . " " . $add_cond : "") .
") ";

if ($where) {
    if (isset($condition['concat']) && strtolower($condition['concat']) == 'or') {
        $concat = " OR ";
    } else {
        $concat = " AND ";
    }
    $where .= $concat . $wh;
} else {
    $where = $wh;
}

```

```

        }
    }
}
}
if (!empty($data['sort'])) {
    $field_sort_info = $this->model_doctype_doctype->getField($data['sort']);
    if ($field_sort_info) {
        $joins[] = "LEFT JOIN " . DB_PREFIX . "field_value_" . $this->db->escape($field_sort_info['type']) . " fv" . $this->db->escape(str_replace("-", "", $data['sort'])) . " ON (fv" . $this->db->escape(str_replace("-", "", $data['sort'])) . ".document_uid=d.document_uid AND fv" . $this->db->escape(str_replace("-", "", $data['sort'])) . ".field_uid = " . $this->db->escape($data['sort']) . ") ";
    }
}
if (!empty($joins)) {
    $joins = array_unique($joins);
    $sql .= implode(" ", $joins);
}
if (!empty($data['draft_less'])) {
    $sql .= "WHERE d.draft < " . (int) $data['draft_less'] . " ";
} else {
    $sql .= "WHERE d.draft < 2 ";
}

if (!empty($where)) {
    $wh = explode(" AND ", $where);
    $sql .= " AND (" . implode(" AND (", $wh) . ") ";
}
if (!empty($data['author_uids'])) {

```



```

        $sql .= "AND d.author_uid IN (" . implode(",", $data['author_uids']) . ")
";
    }
    if (!empty($data['department_uids'])) {
        $sql .= "AND d.department_uid IN (" . implode(",", $data['department_uids']) . ") ";
    }
    if (!empty($data['doctype_uids'])) {
        $sql .= "AND d.doctype_uid IN (" . implode(",", $data['doctype_uids']) . ") ";
    }
    if (!empty($data['document_uids'])) {
        $sql .= "AND d.document_uid IN (" . implode(",", $data['document_uids']) . ") ";
    }
    if (!empty($data['route_uid'])) {
        $sql .= "AND d.route_uid = " . $this->db->escape($data['route_uid']) . " ";
    }
    if (!empty($data['sort']) && $field_sort_info) {
        $query = $this->db->query("SELECT DATA_TYPE FROM
information_schema.COLUMNS WHERE TABLE_SCHEMA=" . $this->db->escape(DB_DATABASE) . " AND TABLE_NAME='field_value_' . $this->db->escape($field_sort_info['type']) . " AND COLUMN_NAME='value'");
        if ($query->row['DATA_TYPE'] == 'datetime' || $query->row['DATA_TYPE'] == 'date' || $query->row['DATA_TYPE'] == 'time' || $query->row['DATA_TYPE'] == 'int' || $query->row['DATA_TYPE'] == 'tinyint' || $query->row['DATA_TYPE'] == 'smallint' || $query->row['DATA_TYPE'] == 'mediumint' || $query->row['DATA_TYPE'] == 'bigint' || $query->row['DATA_TYPE'] == 'decimal' || $query->row['DATA_TYPE'] == 'float' ||

```

```

$query->row['DATA_TYPE'] == 'double' || $query->row['DATA_TYPE'] ==
'real') {
    $sql .= " ORDER BY fv" . $this->db->escape(str_replace("-", "",
$data['sort'])) . ".value " . $this->db->escape(strtoupper($data['order'] ?? "ASC")) .
" ";
} else {
    $sql .= " ORDER BY fv" . $this->db->escape(str_replace("-", "",
$data['sort'])) . ".display_value " . $this->db->escape(strtoupper($data['order'] ??
"ASC")) . " ";
}
}
if (!empty($data['limit']) && isset($data['start'])) {
    $sql .= " LIMIT " . (int) $data['start'] . "," . (int) $data['limit'];
}

$query = $this->db->query($sql);
if (!empty($data['function'])) {
    $result = $query->row['result'];
} else {
    $result = array();
    foreach ($query->rows as $document) {
        $result[] = $document['document_uid'];
    }
}
return $result;
}

```