

Инженерная школа информационных технологий и робототехники

Направление подготовки: 09.03.04 «Программная инженерия»

Отделение информационных технологий

БАКАЛАВРСКАЯ РАБОТА

| Тема работы |
|--|
| Разработка программной системы визуализации многомерных данных на основе кривых Эндрюса |

УДК 004.651.3:004.932.2

Студент

| Группа | ФИО | Подпись | Дата |
|--------|-------------------------------|---------|------|
| 8К61 | Ширькалов Анатолий Максимович | | |

Руководитель ВКР

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|----------------------|-----------------------------|------------------------|---------|------|
| Доцент ОИТ ИШИТР ТПУ | Марухина Ольга Владимировна | к.т.н. | | |

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|----------------------|------------------------|------------------------|---------|------|
| Доцент ОСТН ШБИП ТПУ | Спицына Любовь Юрьевна | к.э.н. | | |

По разделу «Социальная ответственность»

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|---------------------|-----------------------------|------------------------|---------|------|
| Доцент ООД ШБИП ТПУ | Белоенко Елена Владимировна | к.т.н. | | |

ДОПУСТИТЬ К ЗАЩИТЕ:

| Руководитель ООП | ФИО | Ученая степень, звание | Подпись | Дата |
|----------------------|-----------------------------|------------------------|---------|------|
| Доцент ОИТ ИШИТР ТПУ | Чердынцев Евгений Сергеевич | к.т.н. | | |

**Планируемые результаты обучения по направлению 09.03.04
«Программная инженерия»**

| Код результата | Результат обучения (выпускник должен быть готов) |
|-----------------------|---|
| P1 | Применять базовые и специальные естественнонаучные и математические знания в области информатики и вычислительной техники, достаточные для комплексной инженерной деятельности. |
| P2 | Применять базовые и специальные знания в области современных информационных технологий для решения инженерных задач. |
| P3 | Ставить и решать задачи комплексного анализа, связанные с созданием аппаратно-программных средств информационных и автоматизированных систем, с использованием базовых и специальных знаний, современных аналитических методов и моделей. |
| P4 | Разрабатывать программные и аппаратные средства (системы, устройства, блоки, программы, базы данных и т. п.) в соответствии с техническим заданием и с использованием средств автоматизации проектирования. |
| P5 | Проводить теоретические и экспериментальные исследования, включающие поиск и изучение необходимой научно-технической информации, математическое моделирование, проведение эксперимента, анализ и интерпретация полученных данных, в области создания аппаратных и программных средств информационных и автоматизированных систем. |
| P6 | Внедрять, эксплуатировать и обслуживать современные программно-аппаратные комплексы, обеспечивать их высокую эффективность, соблюдать правила охраны здоровья, безопасность труда, выполнять требования по защите окружающей среды. |
| P7 | Использовать базовые и специальные знания в области проектного менеджмента для ведения комплексной инженерной деятельности. |
| P8 | Владеть иностранным языком на уровне, позволяющем работать в иноязычной среде, разрабатывать документацию, презентовать и защищать результаты комплексной инженерной деятельности. |
| P9 | Эффективно работать индивидуально и в качестве члена группы, состоящей из специалистов различных направлений и квалификаций, демонстрировать ответственность за результаты работы и готовность следовать корпоративной культуре организации. |
| P10 | Демонстрировать знания правовых, социальных, экономических и культурных аспектов комплексной инженерной деятельности. |
| P11 | Демонстрировать способность к самостоятельной к самостоятельному обучению в течение всей жизни и непрерывному самосовершенствованию в инженерной профессии. |

Федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники

Направление подготовки: 09.03.04 «Программная инженерия»

Отделение информационных технологий

УТВЕРЖДАЮ:
Руководитель ООП

(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ
на выполнение выпускной квалификационной работы

В форме:

| |
|---------------------|
| бакалаврской работы |
|---------------------|

Студенту:

| Группа | ФИО |
|--------|---------------------------------|
| 8K61 | Ширыкалову Анатолию Максимовичу |

Тема работы:

| | |
|---|------------------------|
| Разработка программной системы визуализации многомерных данных на основе кривых Эндрюса | |
| Утверждена приказом директора (дата, номер) | №59-51/с от 28.02.2020 |

| | |
|--|--|
| Срок сдачи студентом выполненной работы: | |
|--|--|

ТЕХНИЧЕСКОЕ ЗАДАНИЕ

| | |
|---|--|
| Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i> | Объектом проектирования в исследовательской работе является программная система для визуализации данных методом кривых Эндрюса. Особые требования к продукту: наличие графического интерфейса пользователя; наличие авторизации пользователей; наличие возможности добавить на визуализации информацию о вероятностях выброса. |
| Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i> | <ol style="list-style-type: none">1. Анализ предметной области2. Проектирование программной системы3. Разработка программной системы4. Финансовый менеджмент5. Социальная ответственность |

| | |
|---|---|
| Перечень графического материала (с точным указанием обязательных чертежей) | 1. Проектирование системы (диаграммы в нотациях <i>IDEF1X</i> , <i>IDEF3</i> , <i>EPC</i> , <i>UML Sequence</i>) 2. Рисунки, демонстрирующие результаты 3. Диаграмма Ганта |
| Консультанты по разделам выпускной квалификационной работы (с указанием разделов) | |
| Раздел | Консультант |
| Финансовый менеджмент, ресурсоэффективность и ресурсосбережение | Спицына Любовь Юрьевна |
| Социальная ответственность | Белоевко Елена Владимировна |
| Названия разделов, которые должны быть написаны на русском и иностранном языках: | |
| | |

| | |
|---|--|
| Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику | |
|---|--|

Задание выдал руководитель:

| | | | | |
|----------------------|-----------------------------|------------------------|---------|------|
| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
| Доцент ОИТ ИШИТР ТПУ | Марухина Ольга Владимировна | к.т.н. | | |

Задание приняли к исполнению студенты:

| | | | |
|--------|-------------------------------|---------|------|
| Группа | ФИО | Подпись | Дата |
| 8К61 | Ширькалов Анатолий Максимович | | |

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа информационных технологий и робототехники

Направление подготовки: 09.03.04 «Программная инженерия»

Уровень образования бакалавриат

Отделение информационных технологий

Период выполнения осенний/весенний семестр 2019/2020 учебного года

Форма представления работы:

| |
|---------------------|
| бакалаврская работа |
|---------------------|

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН
выполнения выпускной квалификационной работы

| | |
|--|--|
| Срок сдачи студентом выполненной работы: | |
|--|--|

| Дата контроля | Название раздела (модуля) / вид работы (исследования) | Максимальный балл раздела (модуля) |
|---------------|---|---------------------------------------|
| | Анализ предметной области | 25 |
| | Проектирование программной системы | 25 |
| | Разработка программной системы | 20 |
| | Финансовый менеджмент, ресурсоэффективность и ресурсосбережение | 15 |
| | Социальная ответственность | 15 |

Составил преподаватель:

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|-------------------------|--------------------------------|------------------------|---------|------|
| Доцент ОИТ ИШИТР ТПУ | Марухина Ольга Владимировна | к.т.н. | | |

СОГЛАСОВАНО:

| Руководитель ООП | ФИО | Ученая степень, звание | Подпись | Дата |
|-------------------------|--------------------------------|------------------------|---------|------|
| Доцент ОИТ ИШИТР ТПУ | Чердынцев Евгений Сергеевич | к.т.н. | | |

ЗАДАНИЕ ДЛЯ РАЗДЕЛА

«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

| Группа | ФИО |
|--------|---------------------------------|
| 8K61 | Ширыкалову Анатолию Максимовичу |

| Школа | ИШИТР | Отделение школы (НОЦ) | ОИТ |
|---------------------|-------------|---------------------------|--------------------------------|
| Уровень образования | Бакалавриат | Направление/специальность | 09.03.04 Программная инженерия |

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

| | |
|--|--|
| 1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих | Бюджет проекта – не более 1,5 млн. руб., В т.ч. на затраты на оплату труда не более 1 млн. руб. |
| 2. Нормы и нормативы расходования ресурсов | Значение показателя интегральной ресурсоэффективности не менее 4 баллов из 5; Значение интегрального финансового показателя не более 1; |
| 3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования | Коэффициент отчислений на уплату во внебюджетные фонды - 30%; Коэффициент накладных расходов - 16%; |

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

| | |
|--|--|
| 1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения | -Анализ конкурентных технических решений |
| 2. Планирование и формирование бюджета научных исследований | Формирование плана и графика разработки: - определение структуры работ; - определение трудоемкости работ; - разработка графика Ганта. Формирование бюджета затрат на научное исследование: - материальные затраты; - затраты на специальное оборудование; - заработная плата (основная и дополнительная); - отчисления на социальные цели; - накладные расходы. |
| 3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования | - Определение потенциального эффекта исследования |

Перечень графического материала (с точным указанием обязательных чертежей):

| |
|--|
| 1. Оценка конкурентоспособности технических решений |
| 2. Матрица SWOT |
| 3. График проведения и бюджет НИ |
| 4. Оценка ресурсной, финансовой и экономической эффективности НИ |

Дата выдачи задания для раздела по линейному графику

Задание выдал консультант:

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|------------------|---------------|------------------------|---------|------|
| Доцент ОСГН ШБИП | Спицына Л. Ю. | | | |

Задание принял к исполнению студент:

| Группа | ФИО | Подпись | Дата |
|--------|-------------------------------|---------|------|
| 8K61 | Ширыкалов Анатолий Максимович | | |

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

| Группа | ФИО |
|--------|---------------------------------|
| 8K61 | Ширыкалову Анатолию Максимовичу |

| Школа | ИШИТР | Отделение (НОЦ) | ОИТ |
|---------------------|-------------|---------------------------|-----------------------------------|
| Уровень образования | Бакалавриат | Направление/специальность | 09.03.04 Программная инженерия |

Тема ВКР:

| | |
|---|--|
| Разработка программной системы визуализации многомерных данных на основе кривых Эндрюса | |
| Исходные данные к разделу «Социальная ответственность»: | |
| 1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения | Объект исследования – программная система. Область применения – визуализация пользовательских наборов данных. Предполагаемые пользователи – члены исследовательских команд, работающие с многомерными данными. |
| Перечень вопросов, подлежащих исследованию, проектированию и разработке: | |
| 1. Правовые и организационные вопросы обеспечения безопасности: <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. | Трудовой кодекс РФ СанПиН 2.2.2/2.4.1340-03 ТОИ Р-45-084-01 ГОСТ 12.2.032-78 |
| 2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия | <ul style="list-style-type: none"> – отклонение показателей микроклимата в помещении – недостаточная освещенность рабочей зоны – отсутствие или недостаток естественного света – повышенный уровень электромагнитных излучений |
| 3. Экологическая безопасность: | – анализ негативного воздействия на окружающую природную среду: утилизация персональных компьютеров, ноутбуков, оргтехники. |
| 4. Безопасность в чрезвычайных ситуациях: | <ul style="list-style-type: none"> – пожар – землетрясение. |

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

Задание выдал консультант:

| Должность | ФИО | Ученая степень, звание | Подпись | Дата |
|------------------------|--------------------------------|---------------------------------|---------|------|
| Доцент ООД ШБИП ТПУ | Белоевко Елена Владимировна | Кандидат технических наук | | |

Задание принял к исполнению студент:

| Группа | ФИО | Подпись | Дата |
|--------|-------------------------------|---------|------|
| 8К61 | Ширькалов Анатолий Максимович | | |

Реферат

Выпускная квалификационная работа содержит 93 страницы, 38 рисунков, 16 таблиц и 24 источника.

Ключевые слова: кривые Эндрюса, визуализация, многомерные данные, поиск выбросов, программная система, веб-приложение, авторизация пользователей, *OAuth2*.

Объектом исследования является программная система визуализации многомерных данных на основе кривых Эндрюса.

Цель работы – предоставить пользователям, не имеющим навыков программирования, удобный способ визуализировать наборы многомерных данных методом кривых Эндрюса.

В процессе исследования проводилось изучение студентом существующих методов визуализации многомерных данных, методов поиска выбросов в многомерных данных, математического аппарата метода кривых Эндрюса и метода поиска выбросов *LoOP*, с последующей реализацией данных методов на языке программирования *Python*.

В результате исследования была спроектирована и частично разработана программная система, позволяющая пользователям, не имеющим навыков программирования, визуализировать наборы многомерных данных методом кривых Эндрюса с использованием веб-интерфейса.

Экономическая эффективность работы заключается в актуальности поставленной задачи, применению современных программных средств, отсутствии прямых конкурентных решений.

В дальнейшем планируется завершение разработки системы, в том числе информационного портала для управления данными системы, развертка веб-сервера на хостинге для предоставления исследовательской группе научного руководителя доступа к работе с программной системой.

Содержание

| | |
|---|----|
| Содержание | 10 |
| Список терминов, сокращений и условных обозначений | 12 |
| Введение..... | 15 |
| 1 Анализ предметной области | 17 |
| 1.1 Кривые Эндрюса | 17 |
| 1.2 Локальная вероятность выброса..... | 19 |
| 2 Проектирование программной системы | 20 |
| 2.1 Выявление требований к системе..... | 20 |
| 2.1.1 Требования к функциональным характеристикам | 21 |
| 2.1.2 Требования к информационной и программной совместимости..... | 32 |
| 2.2 Проектирование архитектуры системы | 33 |
| 2.2.1 Данные в системе | 33 |
| 2.2.2 Проектирование бизнес-процессов | 38 |
| 3 Разработка программной системы..... | 42 |
| 3.1 Выбор средств реализации | 42 |
| 3.2 Авторизация пользователей в системе | 43 |
| 3.3 Пользовательский интерфейс системы..... | 48 |
| 3.3.1 Интерфейс авторизации..... | 48 |
| 3.3.2 Общий вид основной страницы..... | 50 |
| 3.3.3 Панель основных настроек отображения | 51 |
| 3.3.4 Компонент визуализации | 53 |
| 3.3.5 Дополнительные настройки отображения..... | 56 |
| 3.3.6 Режим интерактивной визуализации | 57 |
| 4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение..... | 60 |
| 4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований | 61 |
| 4.1.1 Потенциальные потребители результатов исследования | 61 |
| 4.1.2 Анализ конкурентных технических решений | 62 |
| 4.1.3 SWOT-анализ..... | 64 |
| 4.2 Планирование научно-исследовательских работ..... | 66 |

| | |
|--|----|
| 4.2.1 Структура работ в рамках научного исследования | 66 |
| 4.2.2 Определение трудоемкости выполнения работ | 67 |
| 4.2.3 Разработка графика проведения научного исследования | 68 |
| 4.3 Бюджет научно-технического исследования (НТИ) | 71 |
| 4.3.1 Расчет материальных затрат НТИ | 71 |
| 4.3.2 Расчет затрат на специальное оборудование для научных (экспериментальных) целей | 72 |
| 4.3.3 Основная заработная плата исполнителей темы | 73 |
| 4.3.4 Дополнительная заработная плата исполнителей темы..... | 74 |
| 4.3.5 Отчисления во внебюджетные фонды (страховые отчисления) | 75 |
| 4.3.6 Накладные расходы | 76 |
| 4.3.7 Формирование бюджета затрат научно-исследовательского проекта..... | 76 |
| 4.4 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования..... | 77 |
| 5 Социальная ответственность..... | 79 |
| 5.1 Правовые и организационные вопросы обеспечения безопасности | 79 |
| 5.1.1 Организационные мероприятия обеспечения безопасности | 79 |
| 5.2 Производственная безопасность | 81 |
| 5.2.1 Анализ опасных и вредных производственных факторов..... | 83 |
| 5.2.2 Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов..... | 84 |
| 5.3 Экологическая безопасность..... | 85 |
| 5.4 Безопасность в чрезвычайных ситуациях..... | 86 |
| 5.4.1 Пожар | 86 |
| 5.4.2 Землетрясение..... | 87 |
| Выводы по разделу..... | 88 |
| Заключение | 89 |
| Список литературы | 91 |

Список терминов, сокращений и условных обозначений

БД (База Данных) – представленная в объективной форме совокупность самостоятельных материалов (статей, расчётов, нормативных актов, судебных решений и иных подобных материалов), систематизированных таким образом, чтобы эти материалы могли быть найдены и обработаны с помощью электронной вычислительной машины (ЭВМ).

Веб-сайт – одна или несколько логически связанных между собой веб-страниц.

Веб-приложение — клиент-серверное приложение, в котором клиент взаимодействует с сервером при помощи браузера, а за сервер отвечает веб-сервер.

Веб-сервер — сервер, принимающий *HTTP*-запросы от клиентов, обычно веб-браузеров, и выдающий им *HTTP*-ответы, как правило, вместе с *HTML*-страницей, изображением, файлом, медиа-поток или другими данными.

HTTP (*HyperText Transfer Protocol*) – протокол прикладного уровня передачи данных (изначально — в виде гипертекстовых документов в формате «*HTML*», в настоящий момент используется для передачи произвольных данных).

HTML (*HyperText Markup Language*) – стандартизированный язык разметки документов во Всемирной паутине.

API (*Application Programming Interface*) – описание способов (набор классов, процедур, функций, структур или констант), которыми одна компьютерная программа может взаимодействовать с другой программой.

SPA (*Single Page Application*) – это веб-приложение или веб-сайт, использующий единственный *HTML*-документ как оболочку для всех веб-страниц и организующий взаимодействие с пользователем через динамически подгружаемые *HTML*, *CSS*, *JavaScript*.

JavaScript (*JS*) – мультипарадигменный язык программирования, поддерживающий объектно-ориентированный, императивный и функциональный стили.

ReactJS, *React* – *JavaScript*-фреймворк с открытым исходным кодом для создания пользовательских интерфейсов.

ООП (объектно-ориентированное программирование) – методология программирования, основанная на представлении программы в виде совокупности объектов, каждый из которых является экземпляром определённого класса, а классы образуют иерархию наследования.

Фреймворк – заготовки, шаблоны для программной платформы, определяющие структуру программной системы; программное обеспечение, облегчающее разработку и объединение разных модулей программного проекта.

Графический интерфейс пользователя (ГИП), графический пользовательский интерфейс (ГПИ) (англ. *graphical user interface*, *GUI*) — система средств для взаимодействия пользователя с компьютером, основанная на представлении всех доступных пользователю системных объектов и функций в виде графических компонентов экрана (окон, значков, меню, кнопок, списков и т. п.).

Drag-and-drop (*D&D*, *DnD*, *DND*, в переводе с английского означает буквально тащи-и-бросай; Бери-и-Брось) — способ оперирования элементами интерфейса в интерфейсах пользователя (как графическим, так и текстовым, где элементы *GUI* реализованы при помощи псевдографики) при помощи манипулятора «мышь» или сенсорного экрана.

IDEF (*I-CAM DEFinition* или *Integrated DEFinition*) – методологии для решения задач моделирования сложных систем, позволяют отображать и анализировать модели деятельности широкого спектра сложных систем в различных разрезах.

DFD (*data flow diagrams*) – методология графического структурного анализа, описывающая внешние по отношению к системе источники и адресаты

данных, логические функции, потоки данных и хранилища данных, к которым осуществляется доступ.

EPC (Event-Driven Process Chain – событийная цепочка процессов) - нотация, используется для описания процессов нижнего уровня. Диаграмма процесса в нотации *EPC*, представляет собой упорядоченную комбинацию событий и функций.

Python — высокоуровневый язык программирования общего назначения, ориентированный на повышение производительности разработчика и читаемости кода. Синтаксис ядра *Python* минималистичен. В то же время стандартная библиотека включает большой объём полезных функций.

MATLAB — пакет прикладных программ для решения задач технических вычислений. Пакет используют более миллиона инженерных и научных работников, он работает на большинстве современных операционных систем.

Введение

Визуализация является важной частью процесса работы с данными, так как позволяет исследователю произвести их качественный анализ. Существует множество различных методов визуализации, однако многие из них становятся малоэффективными при увеличении размерности данных.

Так, например, чтобы визуализировать корреляцию между всеми парами параметров для данных с количеством параметров N , равным 25, необходимо построить

$$\frac{N(N-1)}{2} = 25 \times 12 = 300 \quad (1)$$

диаграмм рассеяния, что затруднительно для восприятия человеком.

Существует несколько методов, позволяющих представить многомерные данные на плоскости. Самыми широко применяемыми среди них являются такие методы как *RadViz*, метод параллельных координат, метод кривых Эндрюса.

Методы представления многомерных данных на плоскости являются мощным инструментом визуализации многопараметрических данных, однако, в то время как методы *RadViz* и метод параллельных координат представлены в некоторых программных системах для визуализации, исследователи, не имеющие опыта программирования, не имеют возможности самостоятельно визуализировать набор многомерных данных методом кривых Эндрюса, так как его реализации присутствуют в основном в библиотеках для языков программирования *Python* и *MATLAB*.

Единственное подобное ПО, не требующее от пользователя написания программного кода для визуализации наборов данных методом кривых Эндрюса, на данный момент является устаревшим, имеющим проблемы с совместимостью с современными ОС, что в совокупности с отсутствием технической поддержки делает его использование крайне затруднительным для исследователей.

Для более детального определения решаемой проблемы был проведен семантический анализ причин. Ниже приведена диаграмма *Fishbone*, отражающая его результаты.



Рисунок 1. Семантический анализ причин (диаграмма *Fishbone*)

Таким образом, для решения проблемы сформулирован следующий список задач:

- а) Изучение литературы и математического аппарата по теме
- б) Выявление требований и составление технического задания
- в) Проектирование системы
- г) Разработка системы
- д) Внедрение системы

Целью данной работы является предоставление пользователям, не имеющим навыков программирования, удобного способа визуализировать наборы многомерных данных методом кривых Эндрюса.

1 Анализ предметной области

1.1 Кривые Эндрюса

Главной особенностью метода визуализации многомерных данных, описанного Эндрюсом в его работе [1], является возможность представить данные любой размерности в виде кривых на плоскости. При этом каждой записи в данных ставится в соответствие функция в виде ряда Фурье:

$$f_x(t) = x_1 2^{-\frac{1}{2}} + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \dots, \quad (2)$$

где $x(x_1, x_2, \dots, x_m)$ – запись (точка) из набора данных,

$x_i (i = 1, \dots, m)$ – изменяемые переменные,

m – размерность данных.

Полученные кривые строятся на плоскости в промежутке $-\pi < t < \pi$ и обладают рядом важных свойств. Во-первых, для любых точек x и y

$$f_x(t) + f_y(t) = f_{x+y}(t), \quad (3)$$

откуда следует, что множество полученных кривых обладает статистическими характеристиками, сходными с таковыми у множества данных.

Среди них: среднее значение (4), расстояния между элементами (5).

$$f_{\bar{x}}(t) = \frac{1}{n} \sum_{i=1}^n f_{x_i}(t), \quad (4)$$

где \bar{x} – средний вектор,

$$\begin{aligned} \|f_x(t) - f_y(t)\|_{L_2} &= \int_{-\pi}^{\pi} [f_x(t) - f_y(t)]^2 dt = \\ &= \pi \sum_{i=1}^m (x_i - y_i)^2 = \pi \|x - y\|^2. \end{aligned} \quad (5)$$

Так же, при условии, что параметры $x_i (i = 1, \dots, m)$ являются некоррелируемыми случайными величинами с дисперсией σ^2 , справедливо

$$\begin{aligned} D[f_x(t)] &= \sigma^2 (2^{-1} + \sin^2(t) + \cos^2(t) + \dots) = \\ &= \begin{cases} \frac{1}{2} \sigma^2 m, & \text{если } m \text{ нечетно,} \\ \frac{1}{2} \sigma^2 \left[m - 1 + 2 \sin^2\left(\frac{mt}{2}\right) \right], & \text{если } m \text{ четно.} \end{cases} \end{aligned} \quad (6)$$

Очевидно, что независимо от значения m выполняется следующее неравенство

$$\frac{1}{2}\sigma^2(m-1) < D[f_x(t)] < \frac{1}{2}\sigma^2(m+1). \quad (7)$$

Эти свойства позволяют, на основе полученного графика, делать следующие предположения относительно данных:

- а) Если несколько кривых находятся близко друг к другу при всех значениях t , то точки данных, соответствующие этим кривым, близки в соответствии с Евклидовой метрикой. Такая группа кривых отображает кластер точек данных. С другой стороны, если кривая визуальнo сильно отличается от большинства других, соответствующая точка, возможно, является выбросом.
- б) Если несколько кривых находятся близко друг к другу при определенных значениях t_i , то точки данных, соответствующие этим кривым, близки в направлениях, описываемых векторами

$$\mathbf{f}_1(t_i) = (2^{-\frac{1}{2}}, \sin(t_i), \cos(t_i), \sin(2t_i), \dots). \quad (8)$$

Это может позволить определить кластеры данных даже с присутствием дополнительных параметров [1].

- в) Если точка данных \mathbf{y} лежит на прямой, соединяющей точки \mathbf{x} и \mathbf{z} , тогда для всех значений t , $f_y(t)$ находится между $f_x(t)$ и $f_z(t)$.

Таким образом, построение кривых Эндрюса позволяет ответить на следующие вопросы о данных:

- а) Содержат ли данные ярко выраженные кластеры?
- б) Содержат ли данные явные выбросы?

Если данные уже разделены на классы:

- а) Чем классы схожи между собой, а чем отличаются?
- б) Имеются ли кластеры данных внутри классов?
- в) Насколько классы однородны?

Для более эффективного применения данного метода визуализации можно произвести предварительную подготовку данных, в которую входит их нормализация, выделение главных компонент, исключение сильно связанных параметров [1, 2].

1.2 Локальная вероятность выброса

Поиск выбросов является важным шагом в процессе анализа данных. Под выбросами подразумеваются данные с аномальными или нетипичными значениями, требующие особого внимания. Поиск выбросов обычно производится на стадии первичного анализа данных для поиска подозрительных записей, которые могут быть удалены или исправлены на этапе отчистки данных, однако также может быть применен на более поздних этапах для поиска новизны или аномальных значений.

Современные методы поиска выбросов являются в основном локальными, основанными на плотности, методами, что значит, что при расчете значения величины, характеризующей степень принадлежности точки данных к выбросам, учитывается лишь часть набора данных, ближайшая к данной точке в многомерном пространстве. Среди этих методов существует метод, называемый «*LoOP*» (*Local outlier probability* – локальная вероятность выброса), впервые описанный в [4], результатом применения которого к набору данных являются вероятности принадлежности каждой из точек данных к выбросам на интервале $[0, 1]$.

Как описано в [4], оценки принадлежности точек данных к выбросам легко интерпретируемы, а также позволяют проводить сравнения как внутри одного, так и среди нескольких наборов данных, что делает метод подходящим для применения в приложениях для конечного пользователя.

Поиск выбросов является важной частью разрабатываемой системы, так как описанный выше метод кривых Эндрюса, используемый для получения визуализации наборов данных, позволяет эффективно представить все записи из набора на одной плоскости, что в совокупности с возможностью добавить в визуализацию информацию о принадлежности точек данных к выбросам, открывает новые возможности для их анализа. В качестве метода получения информации о принадлежности точек данных к выбросам был выбран метод «*LoOP*», как дающий результаты в наиболее удобоинтерпретируемом виде.

2 Проектирование программной системы

2.1 Выявление требований к системе

Программная система предназначена для использования исследовательскими группами, работающими с многомерными данными. Пользуясь программной системой, пользователи должны иметь возможность получать конфигурируемые визуализации наборов данных, загружаемых в систему, на основе кривых Эндрюса.

Метод кривых Эндрюса подразумевает следующую последовательность действий над данными:

- а) Каждой записи в наборе данных ставится в соответствие функция в виде ряда Фурье по формуле (1);
- б) Полученные кривые строятся на плоскости в промежутке $-\pi < t < \pi$;

Для организации работы исследовательской команды с программной системой в ней предусмотрена система ролей и прав, позволяющая руководителю команды ограничивать возможные действия каждого из пользователей в системе.

Работа с настройками визуализации в системе исполнена в виде «тетрадей» (*Notebook*), позволяющих сохранять их между сеансами пользователя. Также подразумевается возможность создания множества различных «тетрадей» для одного набора данных для обеспечения параллельной работы нескольких пользователей.

В системе не предусмотрено создание отдельных информационных пространств, поэтому один экземпляр системы рассчитан на работу с ним одной исследовательской группы.

2.1.1 Требования к функциональным характеристикам

2.1.1.1 Требования к подсистеме работы с пользователями

- В системе должно быть определено две роли пользователей: Администратор и Исследователь.
- **Создание аккаунтов пользователей в системе:**
 - Аккаунт пользователя с ролью Администратор создается при первом запуске системы с указанием электронной почты руководителя исследовательской команды.
 - Аккаунты пользователей с ролью Исследователь добавляются в систему пользователем с ролью Администратор с указанием электронной почты пользователей-исследователей.
 - При добавлении аккаунта пользователя с ролью Исследователь, пользователь с ролью Администратор определяет права добавляемого пользователя. (подробнее о правах пользователей см. подпункт «Права пользователей в системе» пункта 2.1.1.1)
 - После добавления аккаунта пользователя с любой ролью, на адрес электронной почты, указанный при добавлении аккаунта, отправляется письмо, содержащее ссылку на завершение регистрации пользователя в системе.
 - Ссылка на завершение регистрации должна быть рабочей в течение 24 часов с момента отправки письма.
 - В случае, если в течение 24 часов пользователь не воспользовался отправленной ссылкой, аккаунт пользователя удаляется из системы, и для предоставления данному пользователю аккаунта в системе необходимо заново произвести весь процесс добавления аккаунта.
 - Пользователь, перешедший по недействительной ссылке, должен оказаться на странице веб-приложения, уведомляющей пользователя о том, что ссылка уже не действительна.

- В случае, если пользователь воспользовался ссылкой в течение 24 часов с момента отправления письма с ней, он должен оказаться на странице веб-приложения с формой регистрации, позволяющей ввести пароль пользователя и отображаемое в системе имя пользователя.
- Пароли пользователей должны соответствовать следующим правилам:
 - Пароль должен содержать не менее 8 символов,
 - Пароль должен содержать заглавные и строчные буквы, цифры и специальные символы, не менее одного символа каждого из перечисленных видов.
- На странице веб-приложения, содержащей форму регистрации, должны быть приведены правила составления пользовательских паролей, указанные выше.
- При несоответствии пароля, введенного пользователем в форму регистрации, правилам, указанным выше, пользователю должна быть показана надпись, содержащая сведения о том, что введенный пароль не соответствует требованиям программной системы. Процесс регистрации при этом не должен завершаться.
- При соответствии пароля, введенного пользователем в форму регистрации, правилам, указанным выше, а также заполненном поле формы регистрации, содержащего отображаемое в системе имя пользователя, пользователю должна быть показана надпись, уведомляющая пользователя о завершении процесса регистрации.
- При завершении процесса регистрации пользователя, на адрес электронной почты зарегистрированного пользователя и электронный адрес пользователя, обладающего ролью Администратор, должно быть отправлено электронное письмо, содержащее информацию об успешной регистрации пользователя с указанием адреса его электронной почты и отображаемым в системе именем, введенным пользователем в форму регистрации.

- Неавторизированный пользователь при запросе любой страницы веб-приложения (кроме случая, описанного в подпункте «Создание аккаунтов пользователей в системе») должен быть перенаправлен на страницу с формой авторизации.

- **Авторизация пользователей в системе:**

- После успешного создания аккаунта пользователя он должен иметь возможность авторизоваться в системе с использованием адреса электронной почты и пароля, занесенных в систему на этапе добавления аккаунта.
- У пользователя должна быть возможность выбрать вариант авторизации, при котором при закрытии веб-браузера открытие страниц приложения не потребует повторной авторизации (может быть реализовано в виде опции «Запомнить меня» на форме авторизации).
- Авторизированный пользователь должен иметь возможность выйти из системы с любой страницы веб-приложения.
- После нажатия на кнопку выхода из системы пользователь должен быть перенаправлен на страницу с формой авторизации.

- **Изменение информации о пользователях:**

- Авторизованный пользователь должен иметь возможность изменить следующую информацию о себе без подтверждения с помощью электронной почты:
 - Имя пользователя, отображаемое в системе.
- Авторизованный пользователь должен иметь возможность запросить изменение следующей информации:
 - Адрес электронной почты,
 - Пароль.

После запроса пользователя на изменение вышеперечисленных данных на его адрес электронной почты должно быть отправлено письмо, содержащее

ссылку на страницу веб-приложения с формой, позволяющей ему ввести новые данные.

- Ссылка на изменение данных пользователя должна быть действительна в течение одного часа, по истечении времени она должна вести на страницу с информацией об ошибке.

- При изменении пароля должны действовать правила, перечисленные в подразделе «Создание аккаунтов пользователей в системе».

- В случае изменения адреса электронной почты пользователя, на указанный новый адрес должно быть отправлено письмо, содержащее ссылку на подтверждение этого действия.

- Ссылка на подтверждение изменения адреса электронной почты должна быть рабочей в течение одного часа с момента отправки письма, по истечении времени она должна вести на страницу с информацией об ошибке.

- После перехода пользователем по действующей ссылке на подтверждение изменения адреса электронной почты, на новый адрес электронной почты пользователя и электронный адрес Администратора должно быть отправлено электронное письмо, содержащее информацию об успешном изменении адреса электронной почты пользователя с указанием старого и нового адресов его электронной почты и отображаемым в системе именем.

- В случае успешного изменения пароля или адреса электронной почты пользователя все его активные сеансы в системе должны быть завершены.

- **Изменение статуса пользователя:**

- Администратор должен иметь возможность сделать аккаунт любого пользователя с ролью Исследователь «неактивным».

- Все активные сеансы пользователя, чей аккаунт был сделан «неактивным», должны завершиться.

- Пользователю, чей аккаунт был сделан «неактивным», на адрес электронной почты должно быть отправлено письмо, уведомляющее об этом.
- Пользователь, чей аккаунт был сделан «неактивным», не должен иметь возможности войти в систему с использованием данных этого аккаунта.
- Администратор должен иметь возможность сделать «активным» аккаунт любого пользователя, чей аккаунт был сделан «неактивным».
- Пользователю, чей аккаунт был сделан «активным», на адрес электронной почты должно быть отправлено письмо, уведомляющее об этом.
- Пользователь, чей аккаунт был сделан «активным», должен иметь возможность использовать систему таким же образом, как до того, как его аккаунт был сделан «неактивным».
- **Просмотр информации о пользователях:**
 - Всем авторизованным пользователям системы должен быть доступен для просмотра список «активных» пользователей.
 - Записи в списке пользователей, доступном всем авторизованным пользователям, должны содержать следующую информацию:
 - Отображаемое в системе имя пользователя,
 - Адрес электронной почты пользователя.
 - Пользователю, имеющему роль Администратора, должны быть доступны для просмотра права всех пользователей системы.
 - Пользователю, имеющему роль Администратора, должны быть доступны для изменения права всех пользователей системы.
 - Пользователю, имеющему роль Администратора, должны быть доступны для просмотра «неактивные» пользователи системы.
- **Права пользователей в системе:**
 - В системе должны быть определены следующие группы прав:
 - Права на работу с пользователями,

- Права на работу с наборами данных,
- Права на работу с настройками отображения (объектами *Notebook*).
- В каждой группе прав должны быть определены следующие права:
 - Права на создание и удаление;
 - Права на изменение.
- Авторизированный в системе пользователь должен иметь возможность просмотреть свои права в каждой из групп прав на странице веб-приложения, связанной с аккаунтом пользователя.
- Пользователь, являющийся Администратором, должен иметь все права, перечисленные выше.
- Правами в группе прав «Права на работу с пользователями» должен обладать лишь пользователь, являющийся Администратором.
- Пользователь, являющийся Администратором, должен иметь возможность изменения прав всех пользователей системы с ролью «Исследователь».

В последующих пунктах пользователем будет называться пользователь системы, успешно авторизовавшийся в ней.

2.1.1.2 Требования к подсистеме работы с наборами данных

- **Добавление наборов данных в систему:**
 - Пользователь системы, имеющий права на создание и удаление наборов данных, должен иметь возможность добавить в систему набор данных в одном из нижеперечисленных форматов:
 - *CSV (comma separated values)*,
 - *XLS (Excel)*,
 - *XLSX (Excel)*.

Полный перечень форматов, поддерживаемых при загрузке наборов данных, может быть уточнен в дальнейшем посредством составления дополнительного ТЗ.

- При добавлении набора данных в систему, пользователь должен иметь возможность уточнить следующую информацию о наборе данных:
 - Название набора данных в системе,
 - Названия столбцов (параметров) данных,
 - Является ли набор данных приватным (подробнее см. подраздел «Приватные наборы данных»).
- **Изменение информации о наборах данных:**
 - Пользователь, обладающий правами на изменение информации о наборах данных, должен иметь возможность изменить следующую информацию о наборе данных:
 - Название набора данных в системе,
 - Названия столбцов (параметров) данных,
 - Является ли набор данных приватным (только в случае, если пользователь является создателем этого набора данных, подробнее см. подраздел «Приватные наборы данных»).
- **Приватные наборы данных:**
 - Пользователь должен иметь возможность сделать добавленный им в систему набор данных приватным при условии, что в системе нет настроек отображения, созданных на основе данного набора данных другими пользователями.
 - Приватный набор данных должен быть не виден и не доступен для работы пользователям, кроме создателя набора данных и пользователя с ролью Администратор.
 - Пользователь должен иметь возможность сделать добавленный им в систему и сделанный приватным набор данных публичным.
 - Публичный набор данных должен быть виден и доступен для работы всем пользователям системы, имеющим необходимые права.
- **Архивирование наборов данных:**

- Пользователь системы, имеющий права на создание и удаление наборов данных, должен иметь возможность архивировать набор данных.
- Если в системе на момент архивирования набора данных существуют настройки отображения, созданные на его основе, они тоже архивируются (подробнее о работе с настройками отображения см. в пункте 2.1.1.3).
- Набор данных, архивированный пользователем, а также настройки отображения, созданные на его основе, могут быть восстановлены Администратором.
- Набор данных, архивированный пользователем, а также настройки отображения, созданные на его основе, могут быть полностью удалены из системы Администратором.
- **Просмотр информации о наборах данных:**
 - Всем пользователям системы должен быть доступен для просмотра список неархивированных публичных наборов данных.
 - Пользователю системы должен быть доступен список созданных им неархивированных частных наборов данных.
 - Пользователю, имеющему роль Администратора, должен быть доступен для просмотра список всех частных наборов данных пользователей.
 - Пользователю, имеющему роль Администратора, должны быть доступны для изменения все публичные и частные неархивированные наборы данных пользователей.
 - Пользователю, имеющему роль Администратора, должны быть доступны для просмотра все архивированные наборы данных.
 - Списки наборов данных, описанные выше, должны предусматривать возможность выбора элемента списка, с последующим показом следующей информацией о выбранном наборе данных:
 - Название набора данных в системе,
 - Названия столбцов (параметров) данных,
 - Является ли набор данных частным (подробнее см. подпункт «Частные наборы данных»),

- Отображаемое имя пользователя, добавившего набор данных в систему.

2.1.1.3 Требования к подсистеме работы с настройками отображения (тетрадами, объектами *Notebook*)

- **Добавление настроек отображения в систему:**
 - Пользователь системы, имеющий права на создание и удаление настроек отображения, должен иметь возможность добавить в систему настройки отображения для набора данных, доступном ему.
 - При добавлении настроек отображения в систему, пользователь должен иметь возможность уточнить следующую информацию:
 - Название настроек отображения системе,
 - Используемые столбцы (параметры) данных,
 - Являются ли настройки отображения приватными (подробнее см. подпункт «Приватные настройки отображения»).
- **Изменение информации о настройках отображения:**
 - Пользователь, обладающий правами на изменение информации настройках отображения, должен иметь возможность изменить следующую информацию о настройках отображения:
 - Название настроек отображения системе,
 - Используемые столбцы (параметры) данных,
 - Являются ли настройки отображения приватными (только в случае, если пользователь является создателем этих настроек отображения, подробнее см. подпункт «Приватные настройки отображения»).
- **Приватные настройки отображения:**
 - Пользователь должен иметь возможность сделать созданные им настройки отображения приватными.

- Приватные настройки отображения должны быть не видны и не доступны для работы пользователям, кроме их создателя и пользователя с ролью Администратор.
- Пользователь должен иметь возможность сделать созданные им и сделанные приватными настройки отображения публичным.
- Публичные настройки отображения должны быть видны и доступны для работы всем пользователям системы, имеющим необходимые права.
- **Архивирование настроек отображения:**
 - Пользователь системы, имеющий права на создание и удаление настроек отображения, должен иметь возможность архивировать настройки отображения.
 - Настройки отображения, архивированные пользователем, могут быть восстановлены Администратором.
 - Настройки отображения, архивированные пользователем, могут быть полностью удалены из системы Администратором.
- **Просмотр информации о настройках отображения:**
 - Всем пользователям системы должен быть доступен для просмотра список неархивированных публичных настроек отображения.
 - Пользователю системы должен быть доступен список созданных им неархивированных приватных настроек отображения.
 - Пользователю, имеющему роль Администратора, должен быть доступен для просмотра список всех приватных настроек отображения пользователей.
 - Пользователю, имеющему роль Администратора, должны быть доступны для изменения все публичные и приватные неархивированные настройки отображения пользователей.
 - Пользователю, имеющему роль Администратора, должны быть доступны для просмотра все архивированные настройки отображения.

- Списки настроек отображения, описанные выше, должны предусматривать возможность выбора элемента списка, с последующим переходом на экран визуализации данных на основе выбранных настроек.

2.1.1.4 Требования к подсистеме визуализации

- Пользователь должен иметь возможность посмотреть на визуализацию выбранного им набора данных, полученную методом кривых Эндрюса на основе выбранных им настроек отображения.
- Пользователь, имеющий права на изменение настроек отображения, должен иметь возможность изменить настройки отображения, используемые при визуализации, в том числе:
 - Столбец данных, хранящий информацию о принадлежности данных к классам;
 - При условии наличия столбца, выбранного в качестве содержащего информацию о принадлежности данных к классам, отображаемые классы данных;
 - Столбцы, используемые в визуализации;
 - Порядок столбцов, используемых в визуализации;
 - Необходимость стандартизации данных перед преобразованием в кривые;
 - Необходимость расчета вероятностей локального выброса;
 - Количество соседей, учитываемое при расчете вероятностей локального выброса.
- Пользователь должен иметь возможность обновить визуализацию набора данных.
- Пользователь должен иметь возможность выбрать на полученной визуализации две кривых для сравнения их параметров.

- Пользователь должен иметь возможность выбрать кривую на полученной визуализации для просмотра вероятности локального выброса соответствующей записи в данных.
- Пользователь должен иметь возможность выбрать кривую на полученной визуализации для просмотра линейных отклонений значений параметров соответствующей записи в данных от средних значений этих параметров ближайших записей, использованных при расчете вероятности выброса, выраженных в стандартных отклонениях соответствующих параметров этих записей.
- Пользователь, имеющий права на изменение настроек отображения, должен иметь возможность выбрать требуемый ему порог вероятности локального выброса, все записи в данных с вероятностью, превышающей который будут считаться выбросами.
- Пользователь, имеющий права на изменение настроек отображения, должен иметь возможность выбрать один из следующих режимов отображения выбросов:
 - Режим без выбросов,
 - Режим с одними выбросами,
 - Режим со всеми записями,
 - Режим, где визуализация для данных, не являющихся выбросами, расположена рядом с визуализацией данных-выбросов.

2.1.2 Требования к информационной и программной совместимости

Программная система должна предоставлять возможность пользовательской работы через веб-интерфейс с использованием версий веб-браузеров Google Chrome, Mozilla, Opera или Yandex Browser, выпущенных не ранее, чем за 2 года до момента составления настоящего технического задания.

2.2 Проектирование архитектуры системы

2.2.1 Данные в системе

Важный аспект проектирования архитектуры информационной системы – работа с данными, их потоки между компонентами системы.

В системе было выделены следующие сущности (Рисунок 2):

User – сущность, содержащая информацию о пользователе.

Researcher – сущность, расширяющая сущность *User*, содержащая права пользователя.

Administrator – сущность, расширяющая сущность *User*, содержащая указание на то, что пользователь является Администратором.

Dataset – сущность, содержащая информацию о наборах данных, добавленных в систему.

Notebook – сущность, содержащая настройки отображения, влияющие на визуализацию конкретного набора данных.

На рисунках 3-6 приведены диаграммы в нотации *DFD*, включая контекстную диаграмму, диаграмму декомпозиции и несколько диаграмм, описывающих потоки данных важных функциональных частей системы.

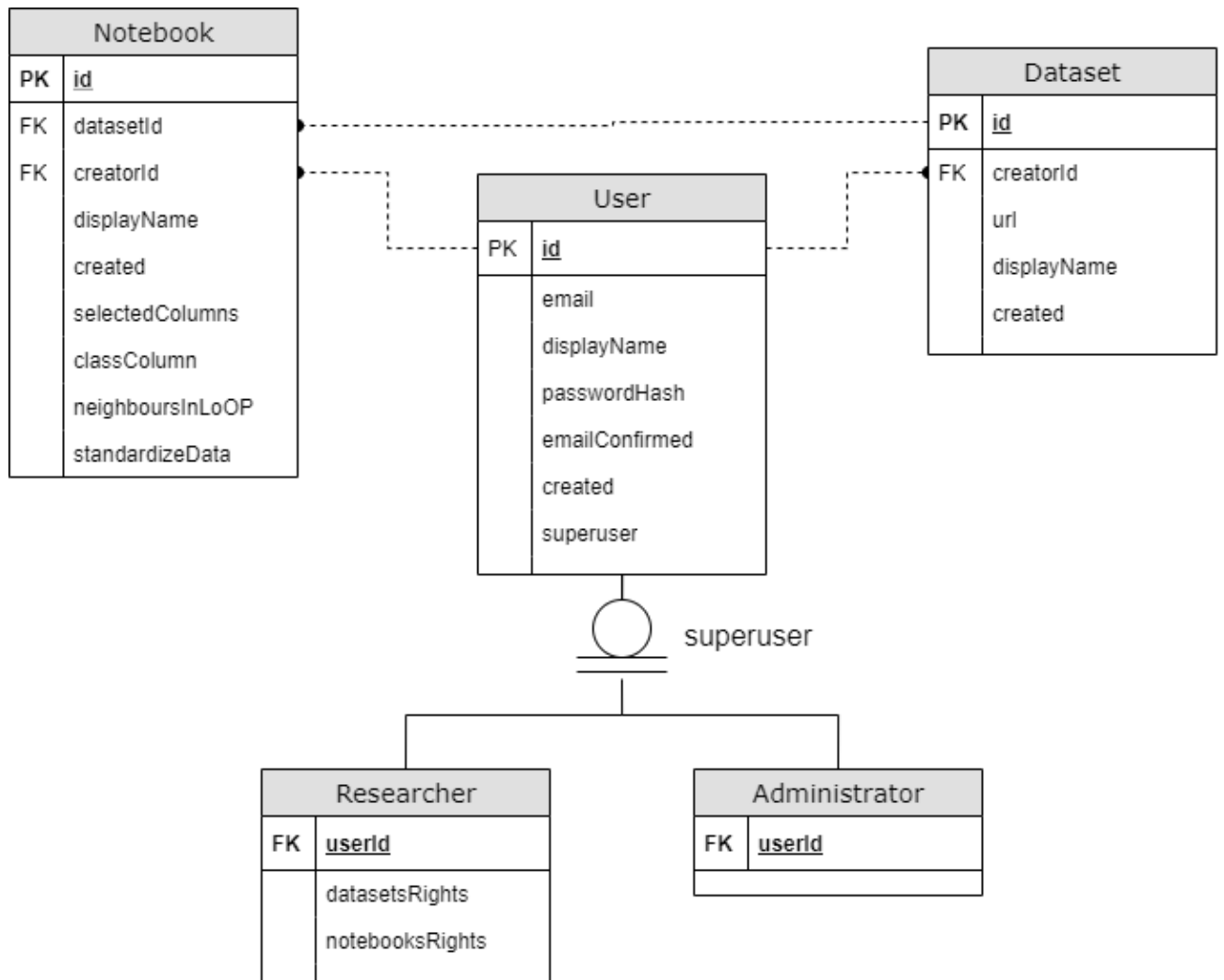


Рисунок 2. Сущности системы и связь между ними
(диаграмма в нотации *IDEF1X*)

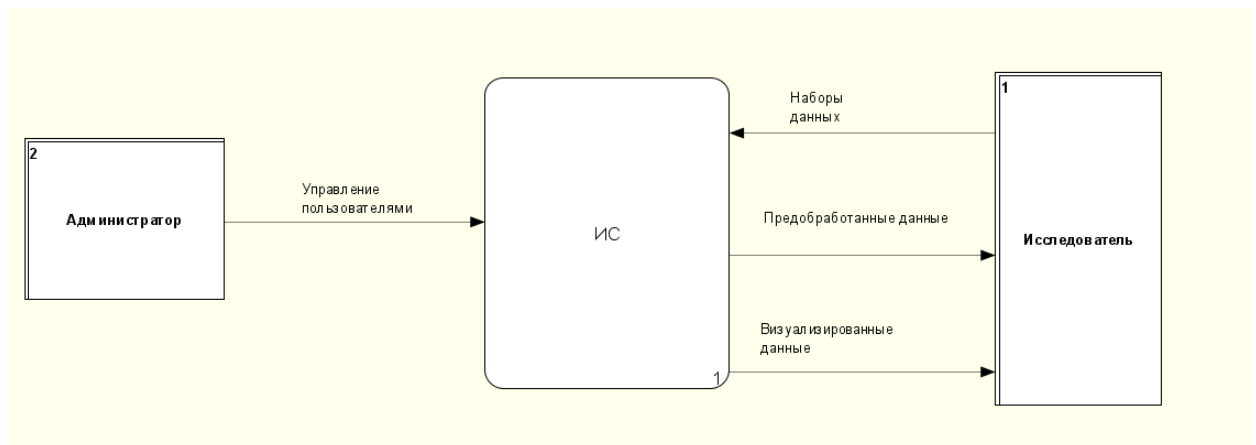


Рисунок 3. Контекстная диаграмма (DFD)

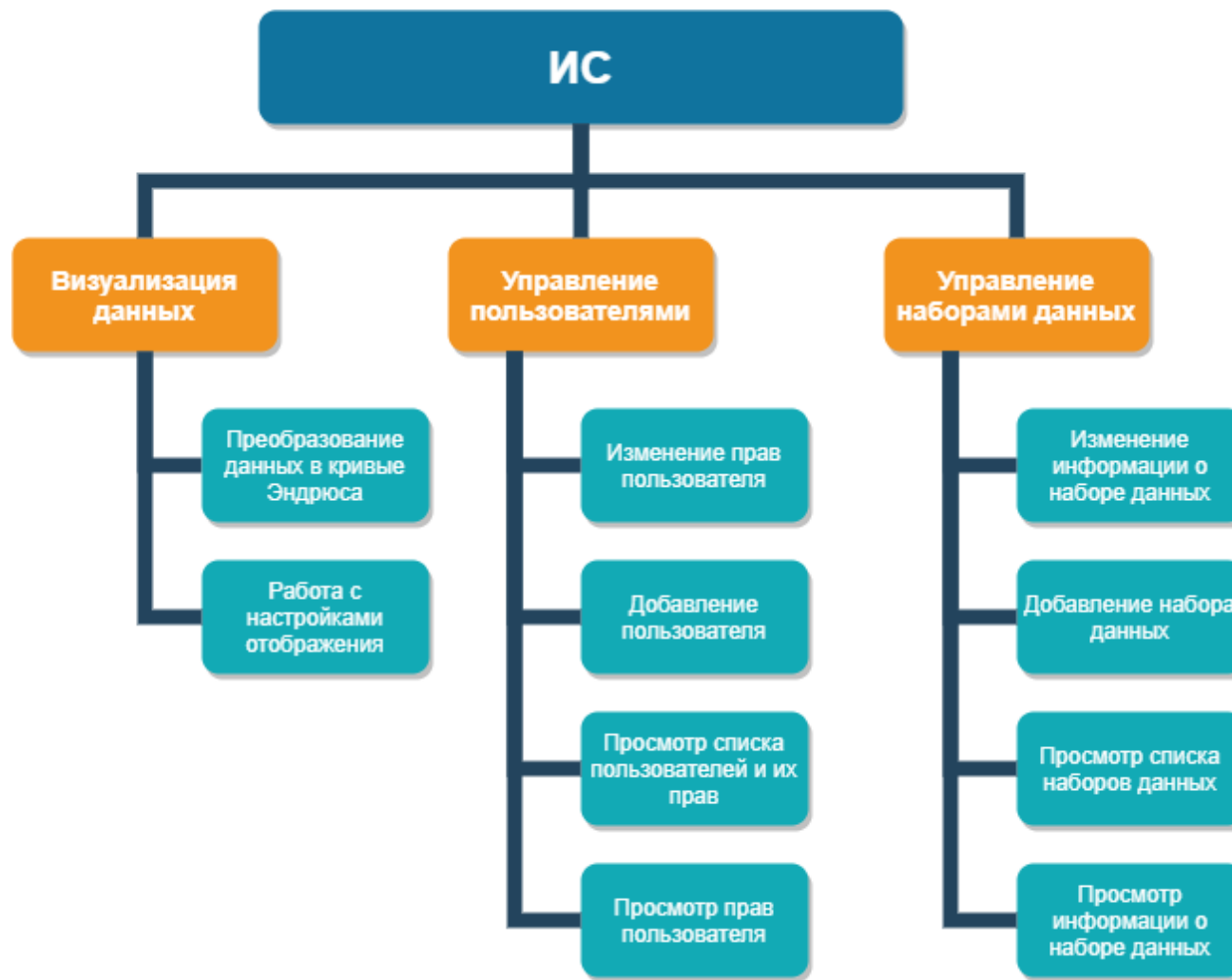


Рисунок 4. Диаграмма декомпозиции (нотация DFD)

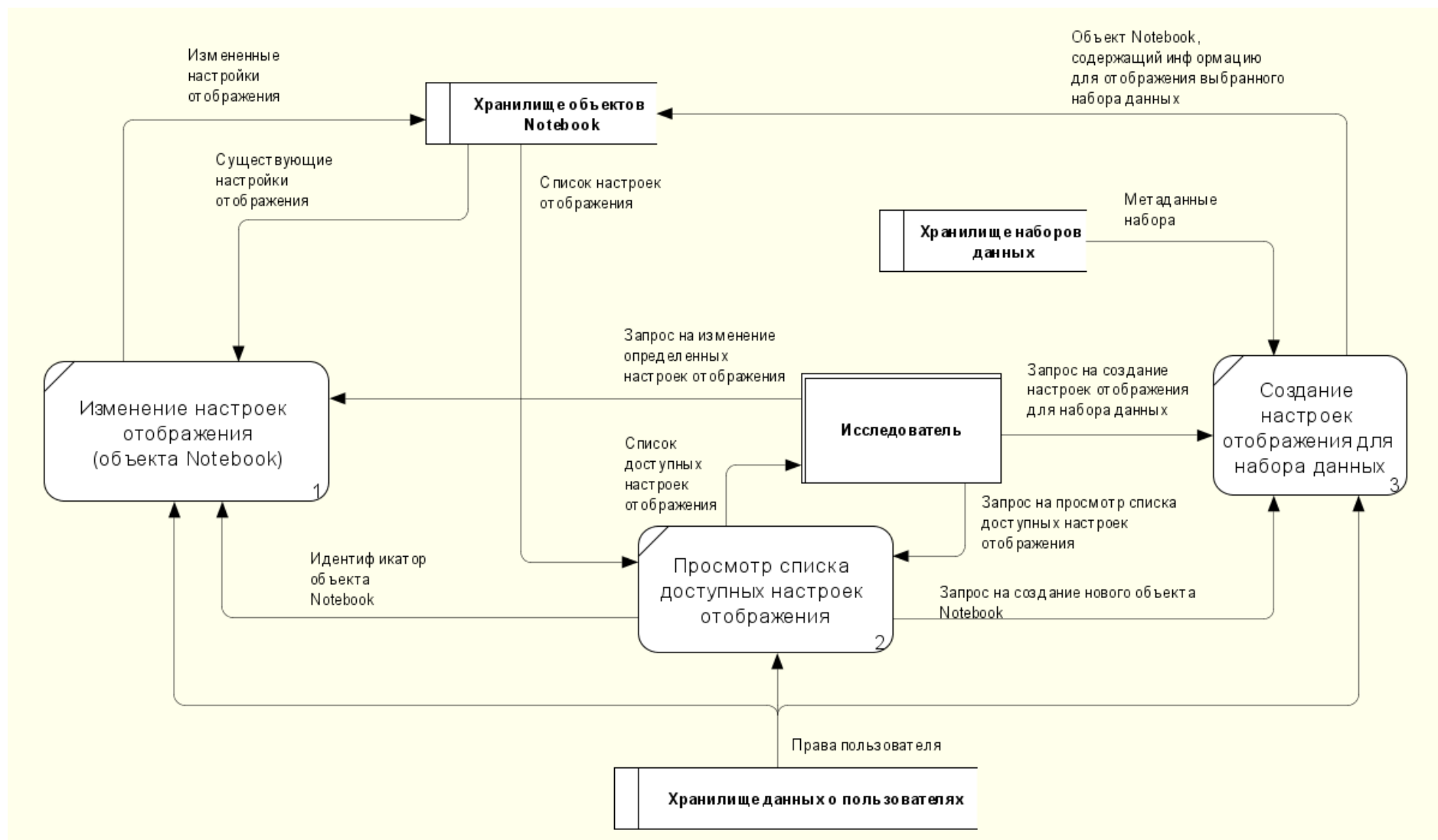


Рисунок 5. DFD-диаграмма декомпозиции работы с настройками отображения

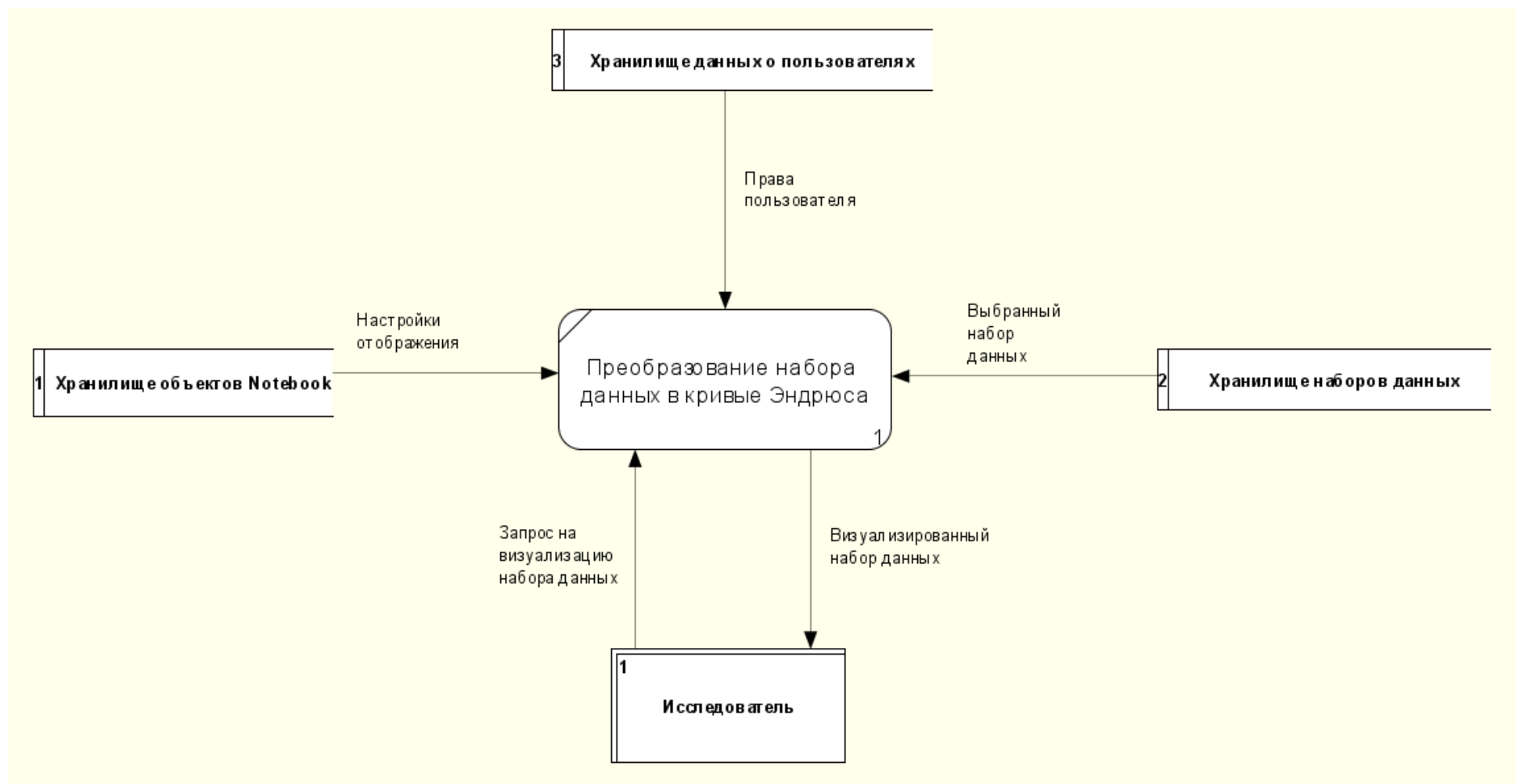


Рисунок 6. DFD-диаграмма процесса преобразования набора данных в кривые Эндрюса

2.2.2 Проектирование бизнес-процессов

При проектировании информационной системы необходимо учитывать события, возникающие в течение бизнес-процесса. Формулирование и четкое понимание этих событий позволяет различить ответственности объектов системы, выявить описать причинно-следственную связь между исходным состоянием и результатом.

Ниже приведены диаграммы в нотации *EPC* (рисунки 7, 8, 9, 10), описывающие процесс визуализации набора данных, полученных исследователем в ходе его работы вне системы, включающий в себя загрузку набора данных в систему, создание настроек визуализации для загруженного набора и построение визуализации набора на основе настроек отображения.

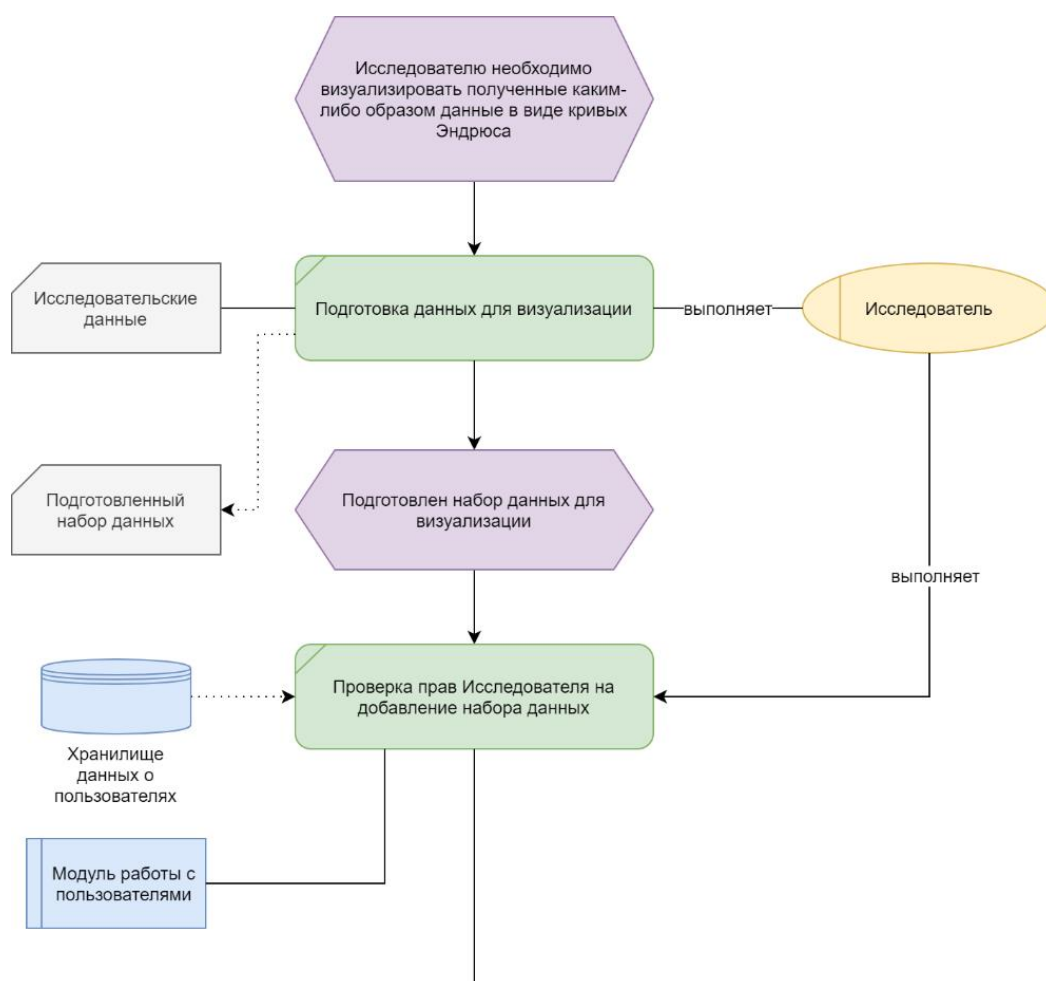


Рисунок 7. Визуализация нового набора данных – первая часть

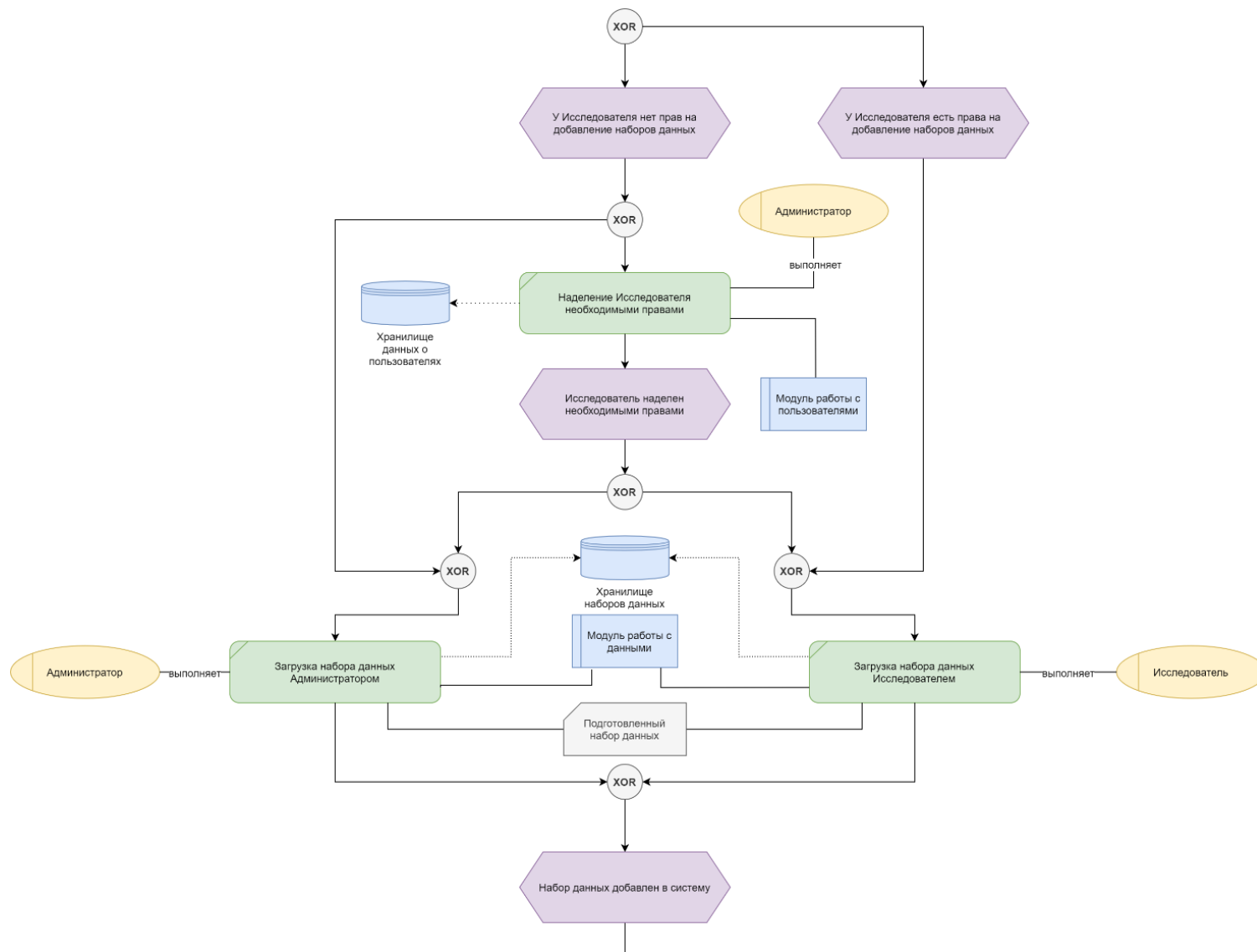


Рисунок 8. Визуализация нового набора данных – вторая часть

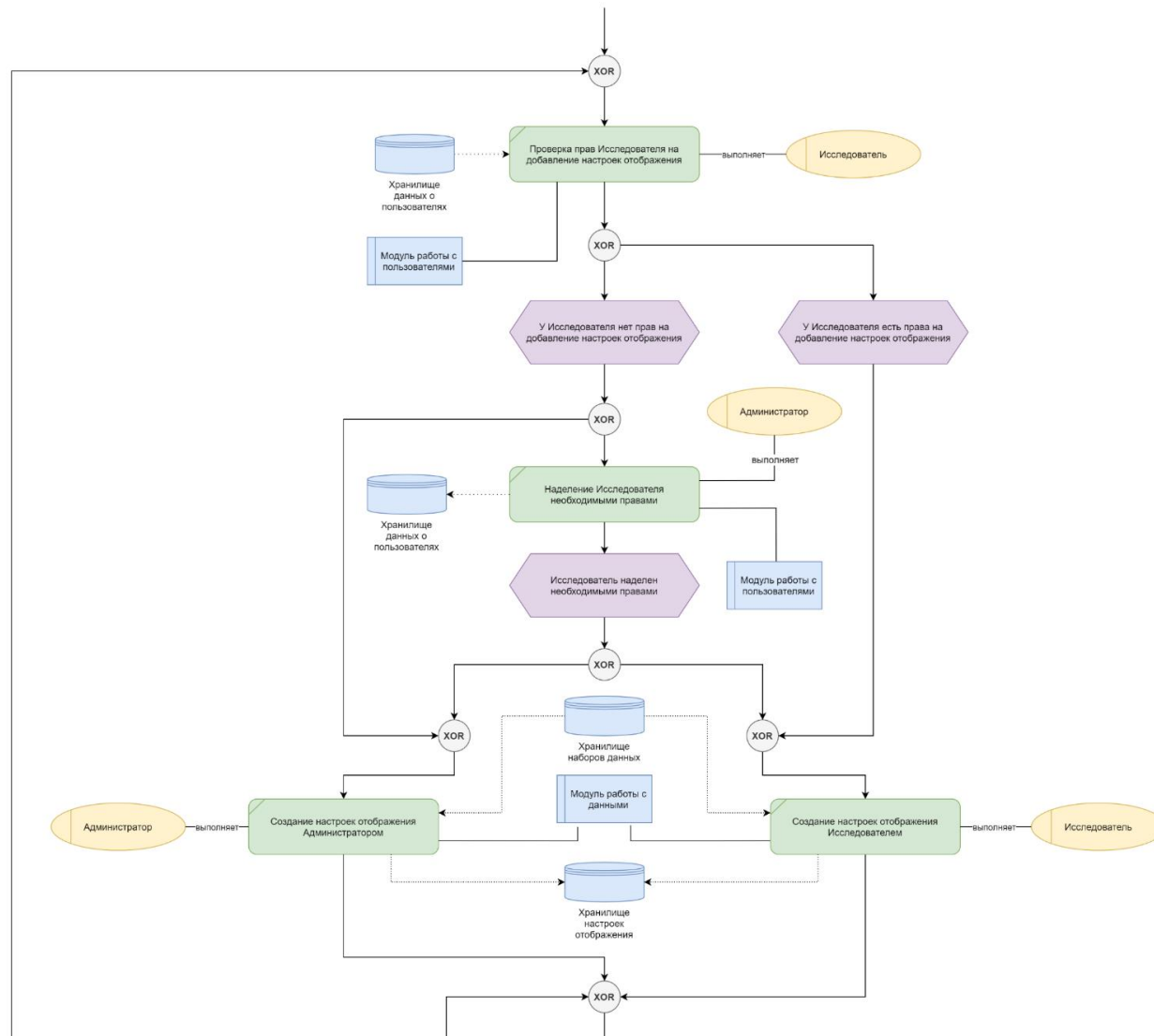


Рисунок 9. Визуализация нового набора данных – третья часть

В качестве архитектуры системы было выбрано веб-серверное приложение с двумя компонентами – веб-сервером и одной веб-страницей, так называемой *SPA (Single Page Application)*, как наиболее подходящее для реализации отзывчивого интерфейса пользователя, в отличие от альтернативного варианта с заготовкой веб-страниц на сервере.

3 Разработка программной системы

3.1 Выбор средств реализации

Для каждого из компонентов системы было проведено сравнение доступных технологий реализации по ряду критериев. Каждому критерию был присвоен вес, затем каждая технология оценивалась по пятибалльной шкале соответствия этому критерию. Ниже, в таблицах 1 и 2 представлены результаты сравнения.

Таблица 1 – Сравнение технологий разработки для веб-сервера

| Критерий | Вес | <i>FastAPI</i> (<i>python</i>) | <i>Spring</i> (<i>Java</i>) | <i>ASP .NET Core</i> (<i>.NET</i>) |
|--|-----|-------------------------------------|----------------------------------|---|
| Опыт разработки | 0.1 | 2 | 1 | 5 |
| Удобство работы с многомерными данными в языке | 0.5 | 5 | 2 | 2 |
| Простота изучения | 0.2 | 5 | 3 | 4 |
| Простота развертывания | 0.2 | 5 | 4 | 3 |
| Итого | 1 | 4.7 | 2.5 | 2.9 |

Как видно из таблицы 1, важнейшим критерием выбора технологии разработки веб-сервера является удобство работы с многомерными данными на языке программирования. В этом отношении язык программирования *python* является наиболее предпочтительным среди рассмотренных вариантов, т.к. обладает множеством библиотек как для работы с данными, так и для создания

безопасных веб-серверов. Таким образом в качестве технологии разработки веб-сервера был выбран фреймворк *FastAPI* для языка программирования *python*.

Таблица 2 – Сравнение технологий разработки *SPA*

| Критерий | Вес | <i>React</i> | <i>Vue</i> | <i>Angular</i> |
|------------------------|-----|--------------|------------|----------------|
| Опыт разработки | 0.5 | 5 | 1 | 2 |
| Простота изучения | 0.3 | 4 | 5 | 4 |
| Простота развертывания | 0.2 | 5 | 5 | 5 |
| Итого | 1 | 4.7 | 3 | 3.2 |

Все вышеперечисленные технологии являются фреймворками для языка программирования *JavaScript*, и в целом имеют схожие возможности и ограничения. Главным критерием при выборе технологии разработки одностраничного приложения было наличие опыта разработки. Таким образом в качестве технологии разработки *SPA* был выбран *React*.

В качестве хранилища данных системы была выбрана база данных *PostgreSQL*, так как она имеет открытый код, подходящую лицензию на использование и большое активное сообщество. Еще одной причиной выбора базы данных *PostgreSQL* стало наличие опыта работы с ней.

3.2 Авторизация пользователей в системе

Авторизация является важной частью большинства программных систем. Она позволяет защитить информационные ресурсы системы от несанкционированного доступа, персонализировать пользовательское взаимодействие с системой. В совокупности с системой прав и ролей, авторизация позволяет повысить уровень безопасности системы, ограничивая

доступ к определенным ресурсам некоторых категорий пользователей.

На данный момент существует несколько общепринятых стандартов организации авторизации в веб-приложениях, однако самым распространенным является стандарт *OAuth2*. В нем определено желательное поведение систем в наиболее распространенных вариантах их архитектуры.

В качестве подхода к организации авторизации пользователей разрабатываемой программной системы был выбран так называемый *Resource Owner Password Credentials Flow* (подход с использованием данных авторизации владельца ресурса, содержащих пароль), имеющий достаточный уровень безопасности для небольших систем.

В веб-серверных приложениях с одной веб-страницей авторизация пользователей в основном осуществляется в виде авторизации *HTTP* запросов к веб-серверу. Одним из вариантов реализации авторизации запросов, допускаемом в вышеописанном подходе, является использование особых токенов, которые сохраняются в хранилище веб-страницы после успешной аутентификации данных, введенных пользователем в поле авторизации. Процесс получения токенов представлен ниже на рисунке 11.

Токены представляют из себя строки особого формата, имеющие в системе определенный срок активности. Информация о времени истечения времени действия токена сохраняется в базу данных после его создания. По истечении этого срока токены не считаются валидными для авторизации.

Предусматривается наличие токенов двух видов:

Access Token или токен доступа используется для авторизации запросов к веб-серверу. При формировании *HTTP* запроса, веб-страница добавляет к нему заголовок *Authorization* со значением «*bearer access_token*», где на место «*access_token*» подставляется полученный ранее токен доступа. Токены доступа имеют относительно короткий «срок жизни», равный 15 минутам.

Refresh Token или токен обновления используется для получения новых токенов доступа и обновления. «Срок жизни» токенов обновления намного больше, чем у токенов доступа, и равняется 15 дням.

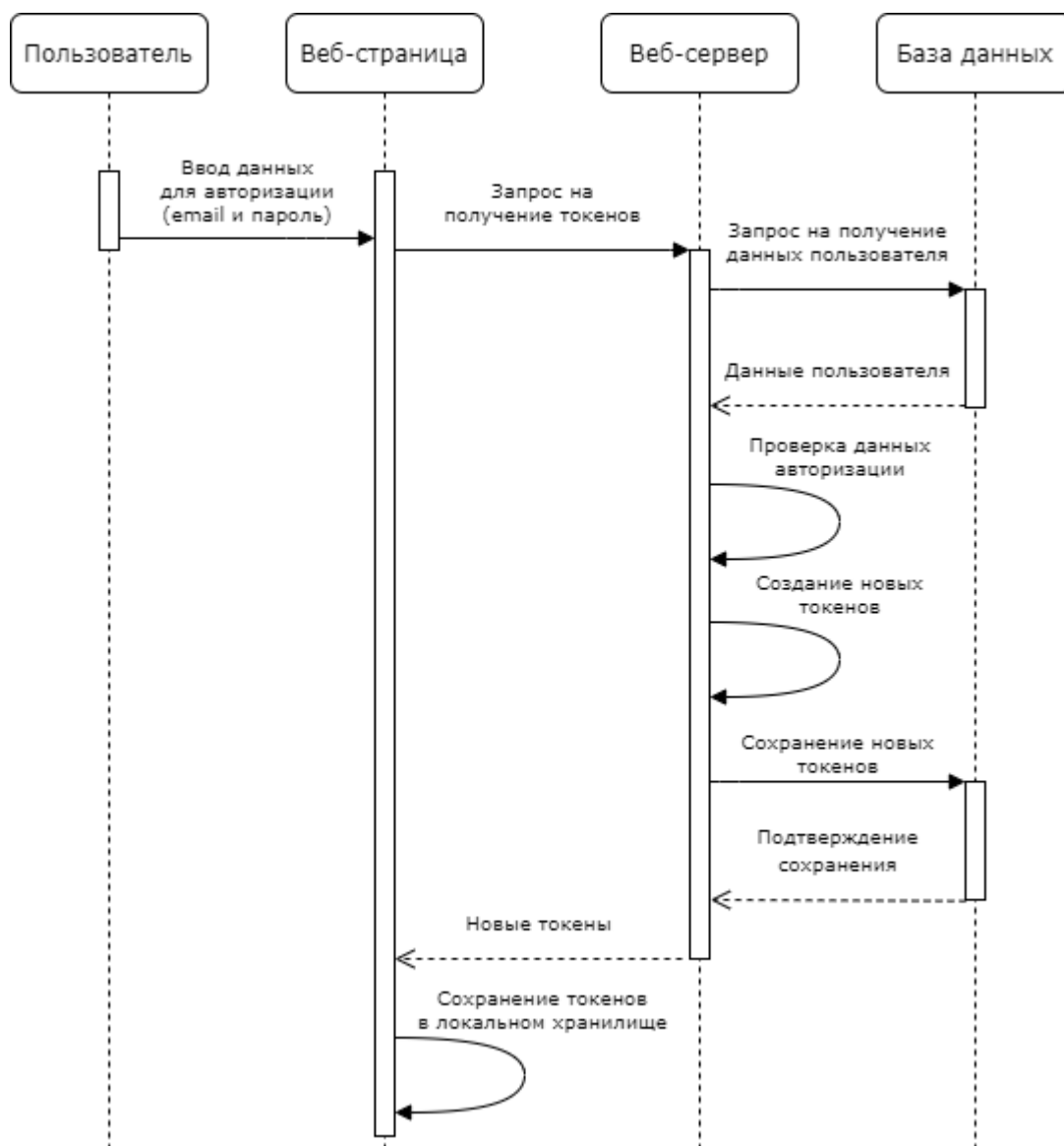


Рисунок 11. Процесс получения токенов в виде диаграммы последовательности в нотации *UML*

В результате процесса, представленного на рисунке 11, в хранилище веб-страницы сохраняется по одному токену каждого вида. Полученный токен доступа используется до тех пор, пока не истечет срок его активности, и в ответ на запрос к веб-серверу не вернется ответ со статус кодом 401 *Unauthorized*. После этого производится попытка обновить токены с использованием токена обновления, и в случае успеха неудавшийся запрос повторяется, а в случае неудачи веб-страница перенаправляется на форму авторизации.

Процесс авторизации HTTP запроса к веб-серверу представлен ниже на рисунках 12-14.

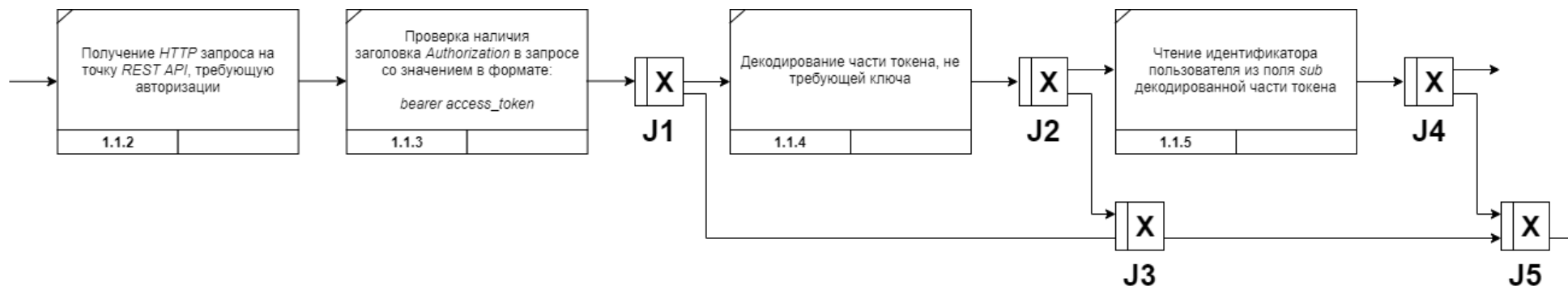


Рисунок 12. Авторизация *HTTP* запроса – первая часть

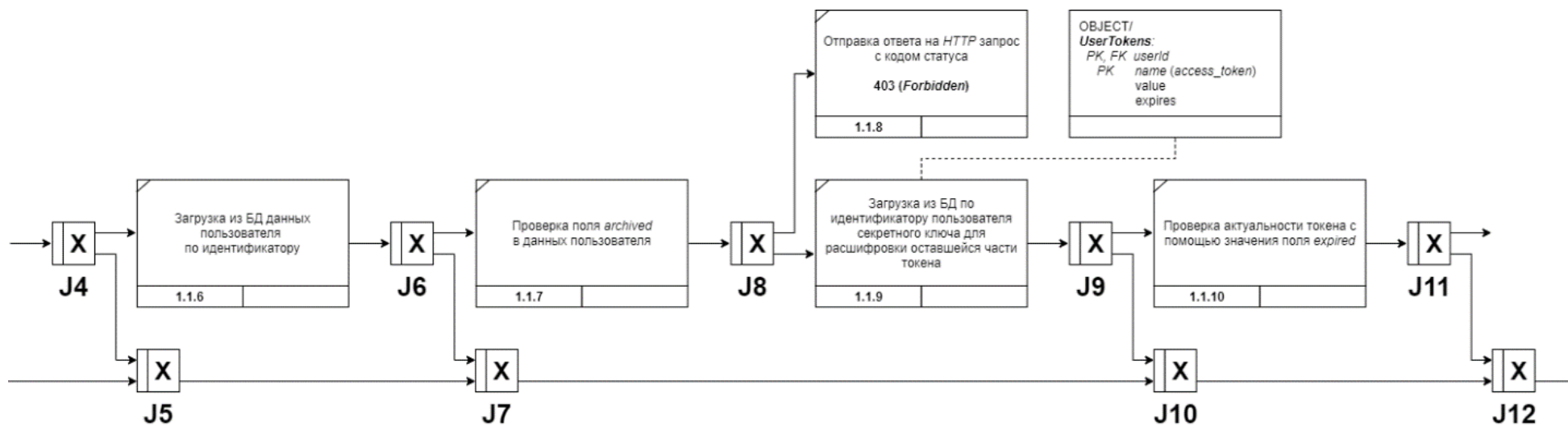


Рисунок 13. Авторизация *HTTP* запроса – вторая часть

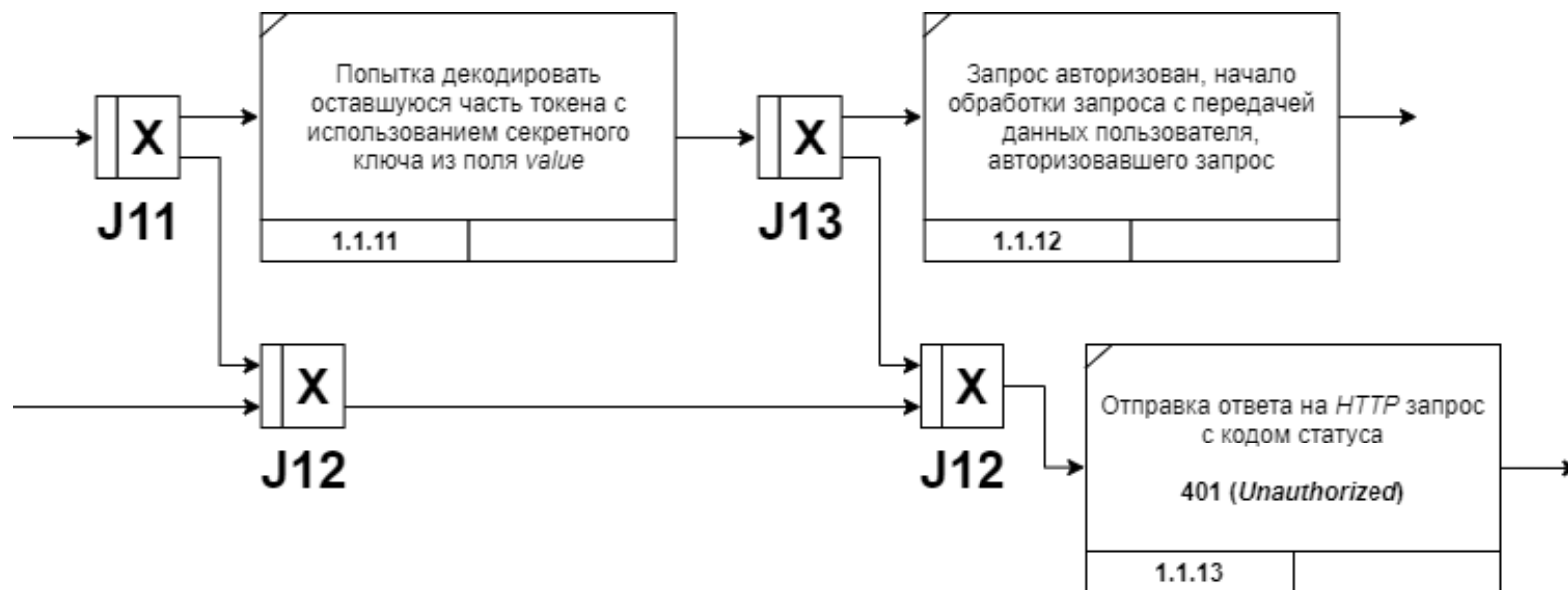


Рисунок 14. Авторизация *HTTP* запроса – третья часть

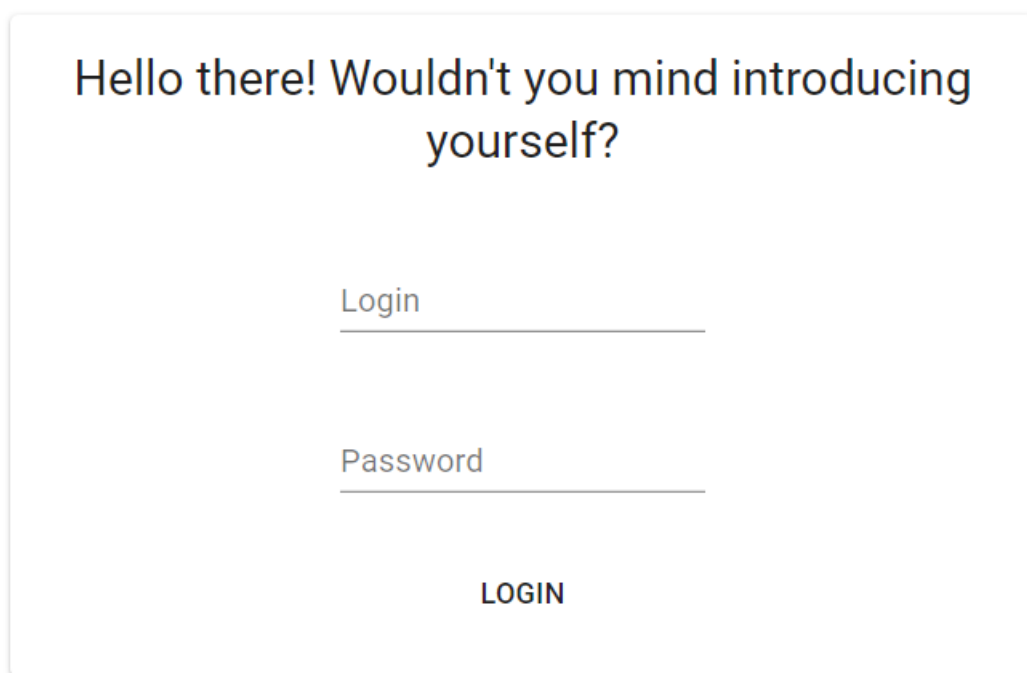
Важно отметить, что не все запросы к веб-серверу требуют обязательной авторизации. Например, стандартом *OAuth2* определено, что на *GET* запрос к веб-серверу по *url*, равному «*базовый_url_сервера/.well-known/openid-configuration*», должен быть возвращен ответ с информацией о схемах авторизации, определенных на сервере.

3.3 Пользовательский интерфейс системы

3.3.1 Интерфейс авторизации

На момент написания данной пояснительной записки разработка системы еще не была завершена. Однако были реализованы основные модули системы, среди которых модуль работы с визуализациями и модуль авторизации. Следующим этапом разработки будет создание портала по управлению данными системы, такими как: пользователи и их права и информация, наборы данных, настройки отображения. Таким образом на данный момент система представляет из себя прототип, предоставляющий возможность оценить основную функциональность системы, без возможности вносить новые данные.

Неавторизованный пользователь, после открытия веб-приложения, увидит перед собой окно авторизации, представленное на рисунке 15. При успешной авторизации пользователь будет перенаправлен к основным экранам приложения, иначе ему будет выведено сообщение об ошибке.



Hello there! Wouldn't you mind introducing yourself?

Login

Password

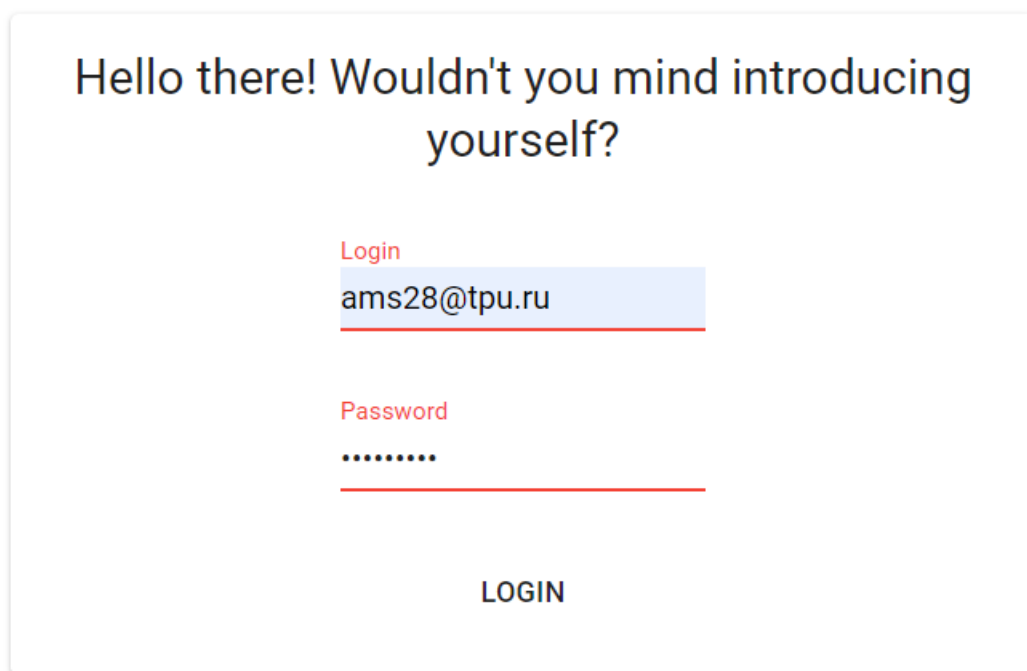
LOGIN

This is a screenshot of a web form for user authentication. It features a light gray background with rounded corners. At the top, a friendly greeting is displayed. Below it, there are two input fields: one for the login name and one for the password. Both fields are currently empty. At the bottom of the form, there is a button labeled 'LOGIN'.

Рисунок 15. Окно авторизации

Данная форма содержит два поля ввода и одну кнопку.

Рисунок 16, представленный ниже, показывает окно авторизации в случае неудачной авторизации.



Hello there! Wouldn't you mind introducing yourself?

Login

ams28@tpu.ru

Password

.....

LOGIN

This is a screenshot of the same web form as in Figure 15, but it shows the state after a failed login attempt. The 'Login' field now contains the email address 'ams28@tpu.ru'. The 'Password' field is filled with eight dots, indicating that the password has been entered but is not visible. The 'LOGIN' button remains at the bottom.

Рисунок 16. Окно авторизации при неверном вводе пользовательских данных

Для привлечения внимания пользователя к возможным ошибкам ввода данных, поля ввода формы авторизации подсвечены красным цветом.

3.3.2 Общий вид основной страницы

В случае успешной авторизации открывается основная страница приложения, представленная на рисунке 17.

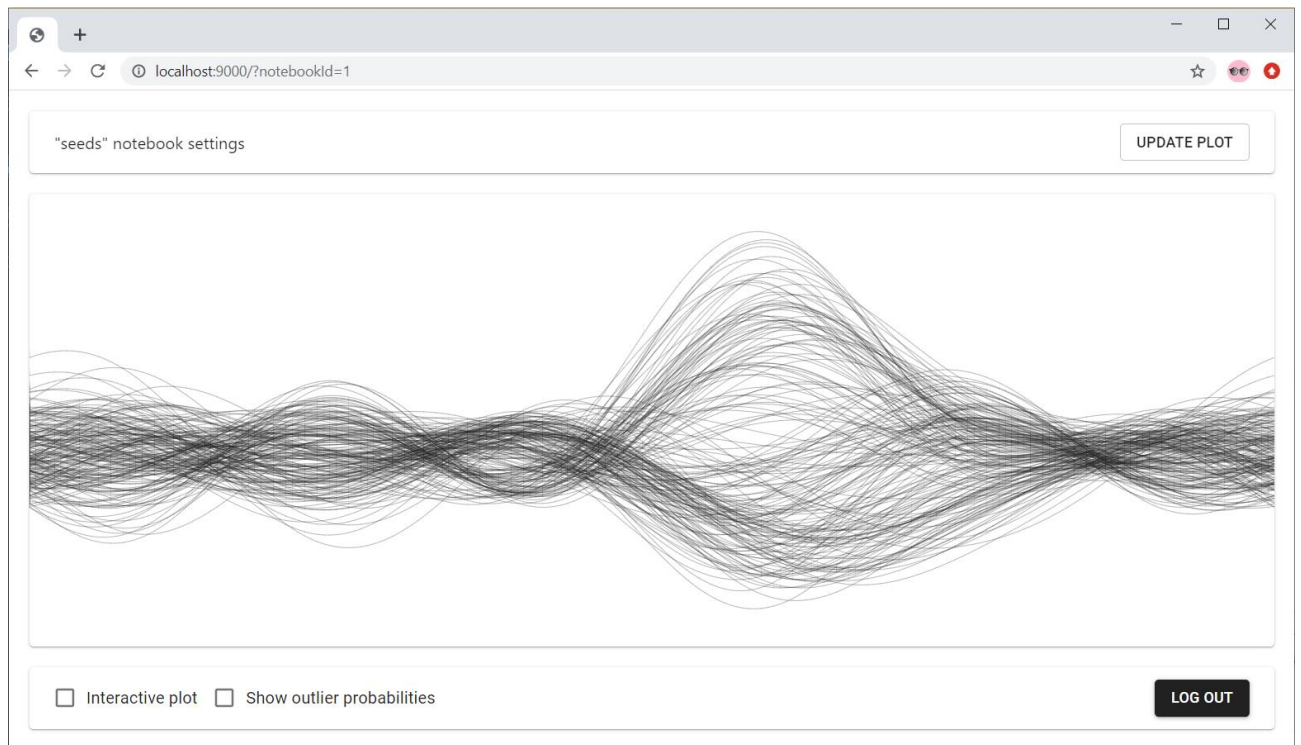


Рисунок 17. Главная страница

На основной странице отображаются следующие элементы экранного интерфейса: свернутая панель основных настроек отображения, компонент с визуализацией, а также компонент с дополнительными настройками отображения.

3.3.3 Панель основных настроек отображения

Клик пользователя на свернутую панель основных настроек отображения открывает ее. Открытая панель основных настроек отображения представлена ниже на рисунке 18.

| Mean | Standard deviation | Minimal value | 25th percentile | 50th percentile | 75th percentile | Maximum value |
|---------|--------------------|---------------|-----------------|-----------------|-----------------|---------------|
| 14.9187 | 2.9200 | 10.5900 | 12.3300 | 14.4300 | 17.4550 | 21.1800 |

Рисунок 18. Открытая панель основных настроек отображения

Как видно из рисунка 18, открытая панель основных настроек отображения состоит из четырех компонентов, расположенных друг над другом:

а) Компонент, содержащий строку заголовок с названием настроек отображения и кнопку для обновления визуализации. Представлен ниже на рисунке 19.

"seeds" notebook settings UPDATE PLOT

Рисунок 19. Компонент открытой панели основных настроек отображения – заголовок

б) Компонент, содержащий элементы управления для изменения следующих настроек отображения: необходимость стандартизации данных, количество соседей, учитываемых при расчете локальной вероятности выброса. Представлен ниже на рисунке 20.

Рисунок 20. Второй компонент открытой панели основных настроек отображения

в) Компонент, содержащий элементы управления для изменения настроек отображения, связанных с параметрами данных. Представлен ниже на рисунке 21. С помощью техники *drag-and-drop*, пользователь имеет возможность выбрать, какие параметры будут участвовать в построении визуализации (колонка «*Selected columns*»), какой параметр набора данных использовать в качестве источника информации о классах (колонка *Class column*, может оставаться пустой при отсутствии такого параметра). Также, при наличии параметра с классами, пользователь имеет возможность просмотреть все возможные значения этого параметра в наборе данных, и выбрать, данные каких классов должны быть отображены на визуализации, а каких нет.

The screenshot shows a user interface with three main sections:

- Selected columns:** A list of parameters that can be selected for visualization: Area, Perimeter, Kernel.Length, Kernel.Width, Asymmetry.Coeff, and Kernel.Groove.
- Other columns:** A list of parameters that are not selected: Compactness.
- Class column:** A section for selecting a parameter to represent classes. It shows 'Type' as the selected parameter. Below it, under 'Class column distinct values', there are three options: 1 (checked), 2, and 3.

Рисунок 21. Компонент открытой панели основных настроек отображения – настройки параметров

г) Компонент, содержащий общую информацию об одном из параметров набора данных, выбираемом пользователем путем наведения курсора на элемент желаемого параметра в компоненте настройки параметров, описанном выше. Представлен на рисунке 22.

| Area | | | | | | |
|---------|--------------------|---------------|-----------------|-----------------|-----------------|---------------|
| Mean | Standard deviation | Minimal value | 25th percentile | 50th percentile | 75th percentile | Maximum value |
| 14.9187 | 2.9200 | 10.5900 | 12.3300 | 14.4300 | 17.4550 | 21.1800 |

Рисунок 22. Компонент открытой панели основных настроек отображения – информация о выбранном параметре

При изменении настроек отображения при помощи элементов управления, описанных выше, не происходит автоматического обновления полученной визуализации. Такое поведение системы выбрано для уменьшения нагрузки на веб-сервер, так как расчет данных для построения кривых на визуализации является довольно ресурсоемкой операцией. Для обновления визуализации пользователю необходимо нажать на кнопку «*Update Plot*», расположенную в правом верхнем углу открытой панели основных настроек отображения. Для удобства, после изменения настроек отображения, кнопка выделяется цветом, как показано на рисунке 23.



Рисунок 23. Вид кнопки «*Update Plot*» до и после изменения настроек отображения

3.3.4 Компонент визуализации

Главным элементом основной страницы приложения является компонент визуализации. На нем отображается визуализация, полученная из выбранного набора данных на основе настроек отображения. При начальной загрузке и обновлении данных для построения кривых в компоненте отображается анимированный элемент загрузки, показанный на рисунке 24.

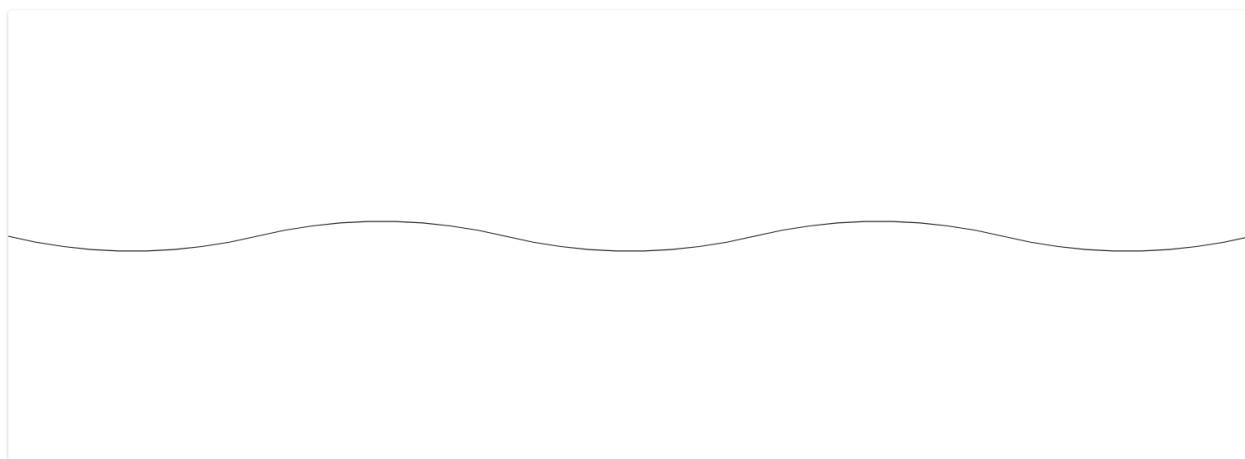


Рисунок 24. Компонент визуализации – элемент загрузки

После загрузки данных для построения кривых, в компоненте отображается визуализация, состоящая из кривых, построенных по полученным данным. Пример отображаемой визуализации представлен на рисунке 25.

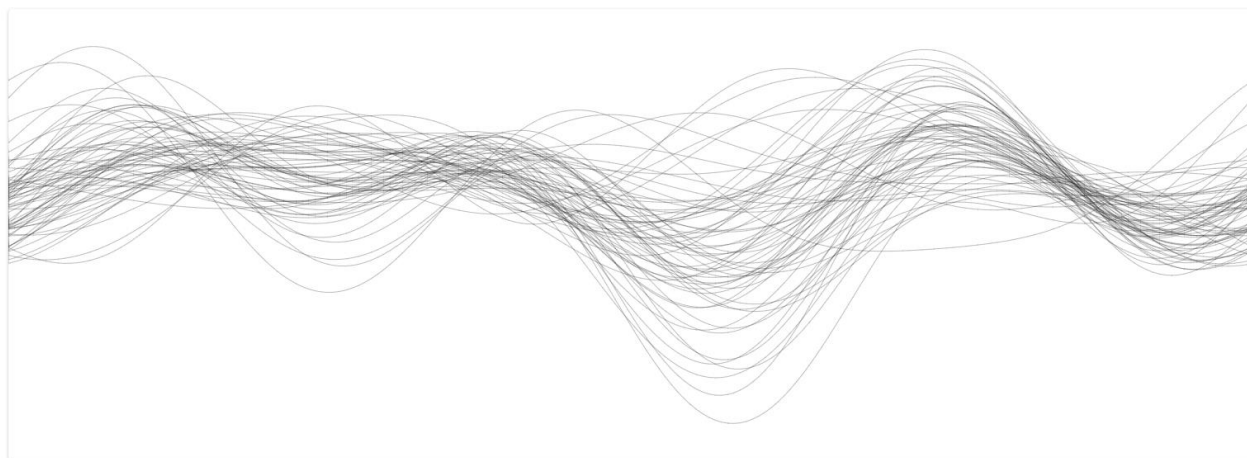


Рисунок 25. Компонент визуализации – пример визуализации

С помощью элемента управления, расположенного в компоненте, описанном ниже в пункте 3.3.5, пользователь имеет возможность включить отображение вероятностей выброса на визуализации. Вид компонента визуализации при включенном отображении вероятностей выброса представлен на рисунке 26.

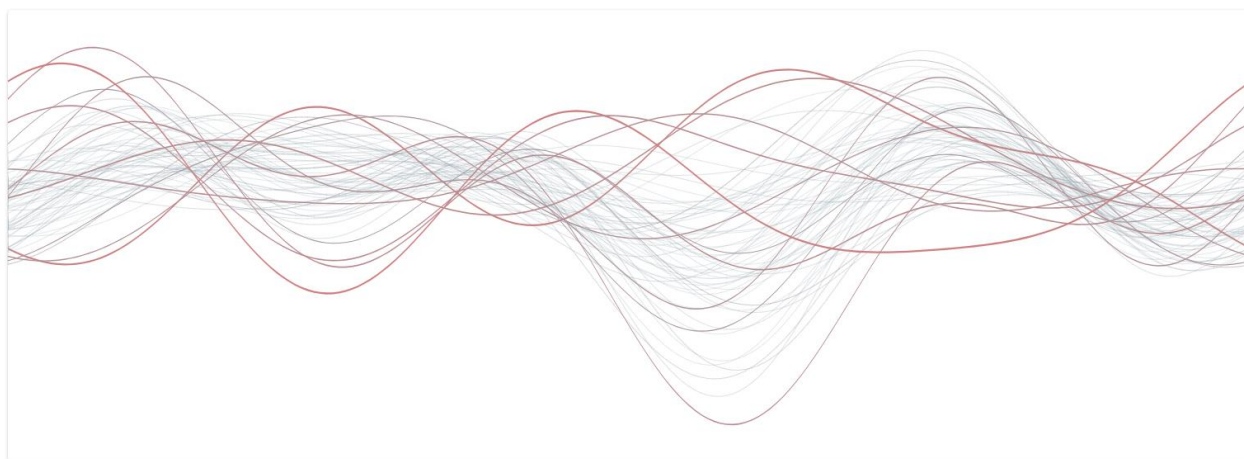


Рисунок 26. Компонент визуализации при включенном
отображении выбросов

Степень перехода цвета линии от светло-серого к красному прямо пропорциональна вероятности того, что точка данных, для которой построена кривая, является выбросом.

При включенном отображении вероятностей выброса пользователь имеет возможность выбрать один из нескольких вариантов отображения выбросов, представленных на рисунках 27 – 30, с помощью элемента управления, описанного в пункте 3.3.5.

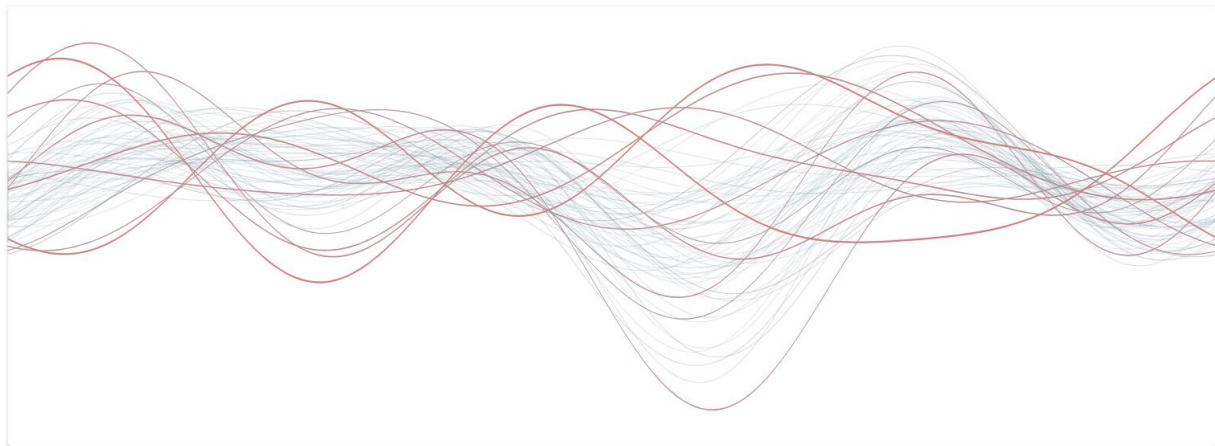


Рисунок 27. Компонент визуализации – вариант «*Combine*»

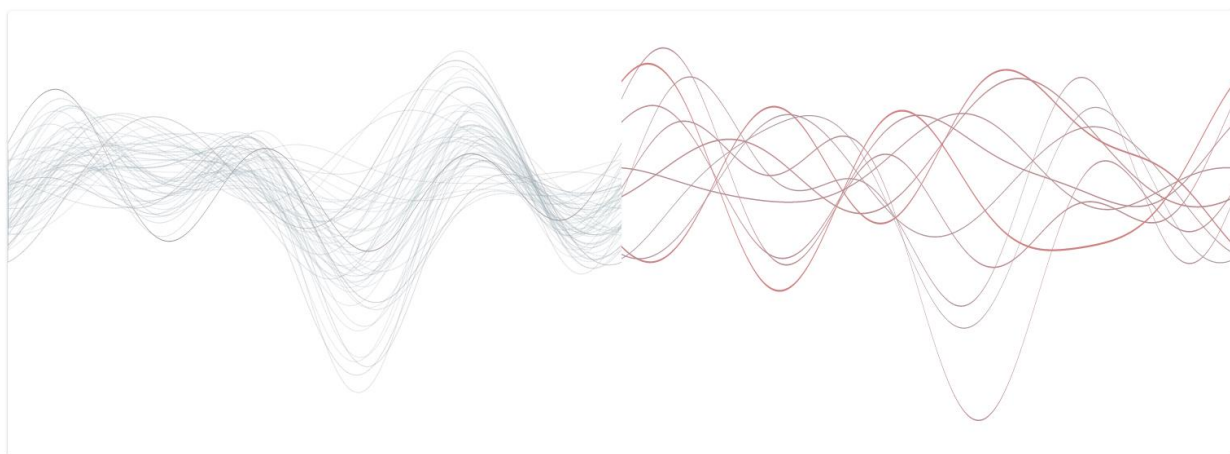


Рисунок 28. Компонент визуализации – вариант «*Split*»

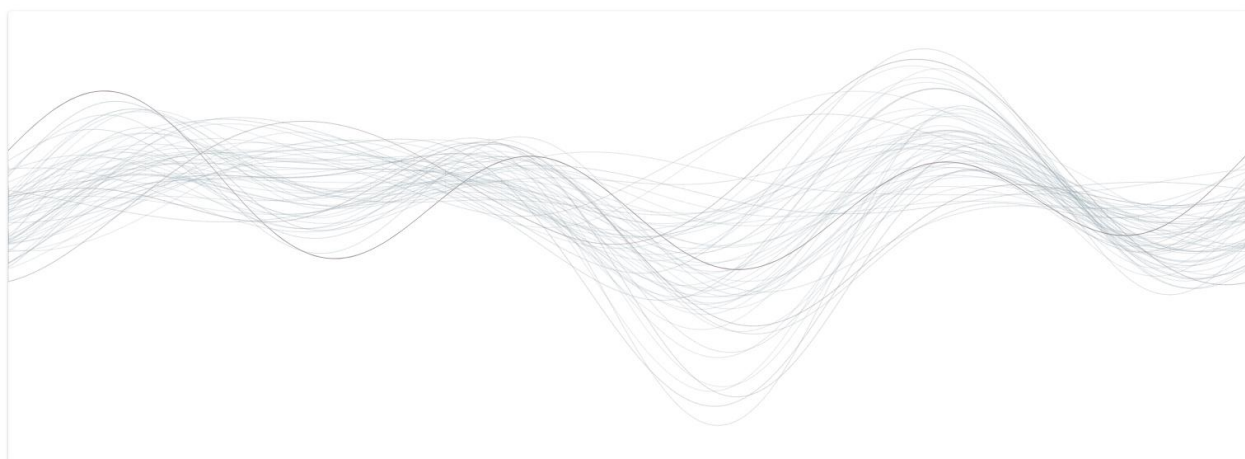


Рисунок 29. Компонент визуализации – вариант «*Hide*»

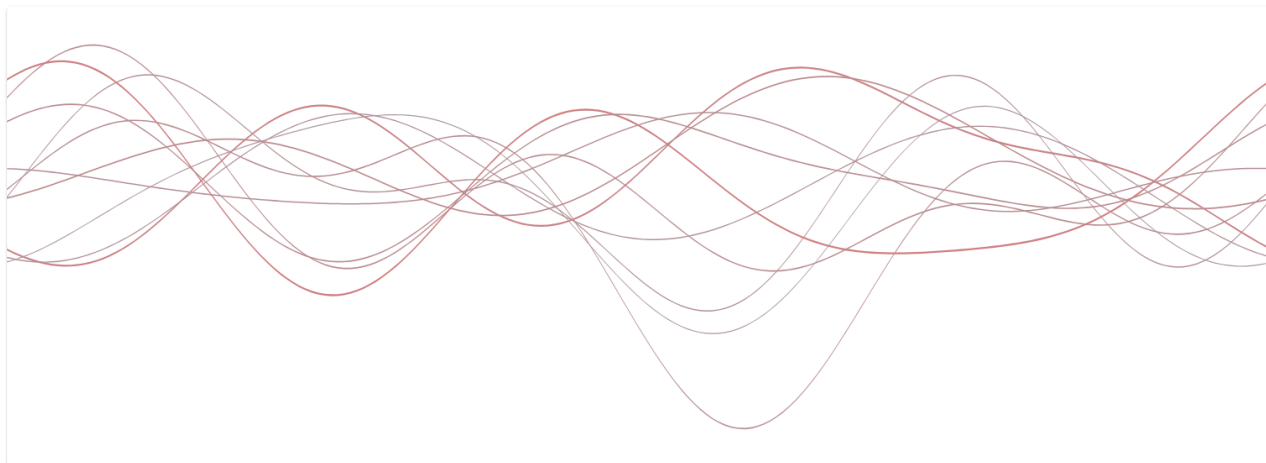


Рисунок 30. Компонент визуализации – вариант «*Outliers only*»

3.3.5 Дополнительные настройки отображения

В нижней части основной страницы расположен компонент, содержащий дополнительные настройки отображения. Его общий вид представлен на рисунке 31.

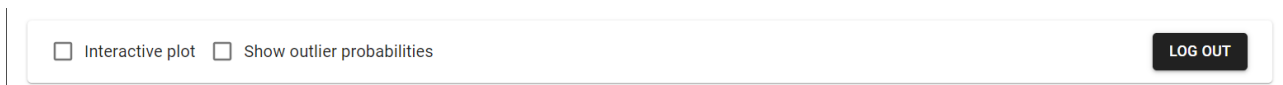


Рисунок 31. Компонент дополнительных настроек отображения

В правой части компонента дополнительных настроек отображения расположена кнопка, при нажатии на которую происходит выход пользователя из системы.

В левой части компонента расположены элементы управления дополнительными настройками отображения. Первый элемент управления, показанный на рисунках 32 и 33, отвечает за включение и отключение режима интерактивной визуализации.

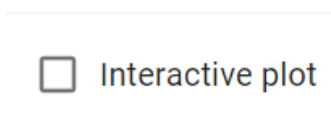


Рисунок 32. Элемент управления – режим интерактивной визуализации (выкл.)

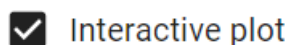


Рисунок 33. Элемент управления – режим интерактивной визуализации (вкл.)

Подробнее о режиме интерактивной визуализации см. пункт 3.3.6.

Вторым элементом управления, расположенном в компоненте дополнительных настроек отображения является элемент, отвечающий за включение и отключение отображения вероятностей выброса на визуализации. Внешний вид элемента показан на рисунках 34 и 35. При включении отображения вероятностей выброса на визуализации в компонент дополнительных настроек отображения добавляются элементы управления, связанные с отображением выбросов на визуализации, как показано на рисунке 35.



Рисунок 34. Элемент управления – отображение вероятностей выброса (выкл.)

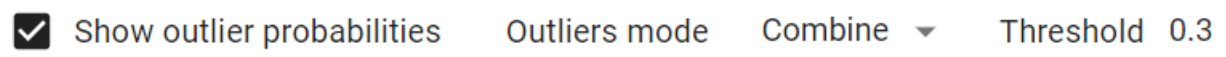


Рисунок 35. Элемент управления – отображение вероятностей выброса (вкл.)

Элемент управления «*Outliers mode*» позволяет выбрать режим отображения выбросов на визуализации, а «*Threshold*» – порог значения вероятности выброса. Кривые для точек данных, для которых значение вероятности выброса превышает порог, будут считаться выбросами и соответственно отображаться в различных режимах отображения выбросов.

3.3.6 Режим интерактивной визуализации

Режим интерактивной визуализации позволяет исследователю, выбрав кривую на визуализации, просмотреть значения параметров соответствующей

точки данных, участвующих в построении визуализации. Работа режима интерактивной визуализации отличается в зависимости от того, включен ли режим отображения вероятностей выброса.

В режиме интерактивной визуализации при выключенном отображении вероятностей выброса пользователь может выбрать на визуализации одну кривую посредством наведения на нее курсора и клика левой клавишей мыши, вторую посредством наведения курсора. При этом кривая, выбранная кликом мыши, окрашивается в ярко-красный цвет, а выбранная наведением курсора – в черный. Также у выбранной одним из способов кривых увеличивается толщина.

На рисунке 36 представлен вид основной страницы при включенном режиме интерактивной визуализации с выключенным отображением вероятностей выброса после выбора пользователем двух кривых. Значения параметров кривой, выбранной кликом, показаны в столбце «*Clicked Line*», а кривой, выбранной наведением курсора – в столбце «*Hovered Line*».

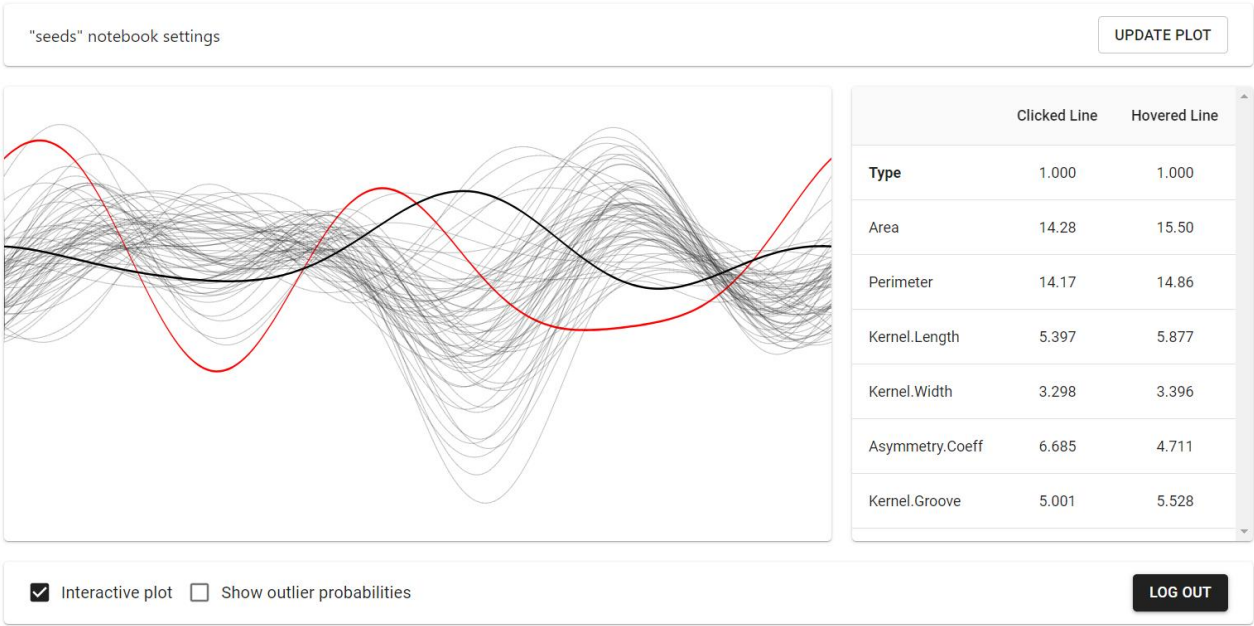


Рисунок 36. Интерактивная визуализация без выбросов

В случае, если отображение вероятностей выброса включено, в режиме интерактивной визуализации пользователь одновременно может просматривать значения параметров только одной точки данных, соответствующей кривой, выбранной кликом по ней. При этом в таблице, кроме значений параметров,

отображаются линейные отклонения значений параметров соответствующей записи в данных от средних значений этих параметров ближайших записей, использованных при расчете вероятности выброса, выраженных в стандартных отклонениях соответствующих параметров этих записей.

На рисунке 37 представлен вид основной страницы при включенном режиме интерактивной визуализации с включенным отображением вероятностей выброса после выбора пользователем кривой. Значения параметров кривой, выбранной кликом, показаны в столбце «*Clicked Line*», а в столбце «*Deviation, σ* » отображаются линейные отклонения точки данных. Также в таблице в строчке «LoOP» имеется информация о значении локальной вероятности выброса для соответствующей точки данных.

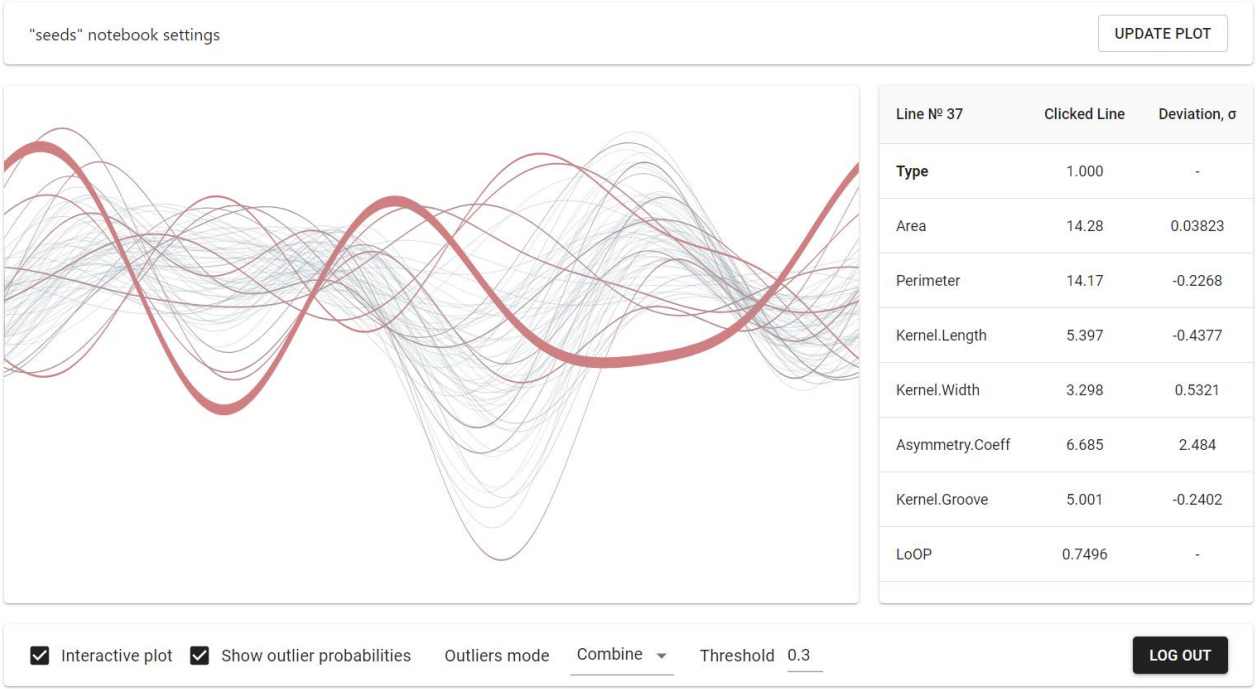


Рисунок 37. Интерактивная визуализация с выбросами

4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

В данном разделе дается оценка коммерческого потенциала и перспективности применения потребителями программной системы визуализации многомерных данных на основе кривых Эндрюса, разработанной в дипломном проекте.

Данная работа направлена на разработку программной системы, предназначенной для предоставления исследователям, не имеющим опыта программирования, возможности визуализировать наборы данных методом кривых Эндрюса. Для достижения поставленных целей необходимо было выполнить ряд задач, в том числе дать оценку коммерческих возможностей проведенного исследования, выявить его ресурсосберегающий потенциал, определить финансовую эффективность исследования.

Целью данного раздела является определение перспективности научно-исследовательского проекта. Задачами раздела являются:

- оценка коммерческого потенциала и перспективности проведения научных исследований с позиции ресурсоэффективности и ресурсосбережения;
- планирование работ по научно-исследовательскому проекту с использованием линейного графика;
- расчет бюджета научного-технического исследования;
- определение экономической эффективности исследования.

4.1 Оценка коммерческого потенциала и перспективности проведения научных исследований

4.1.1 Потенциальные потребители результатов исследования

Для анализа потенциальных потребителей необходимо провести анализ целевого рынка и его сегментирование.

Целевым рынком данного исследования являются исследовательские команды, в работе которых встречается необходимость работы с данными. Можно выделить следующие критерии сегментирования рынка программного обеспечения, позволяющего визуализировать наборы многомерных данных методом Эндрюса: наличие в исследовательской команде опыта программирования и тип программного обеспечения. На основе выбранных данных построим карту сегментирования (таблица 3).

Таблица 3 – Карта сегментирования рынка программного обеспечения, позволяющего визуализировать наборы многомерных данных методом Эндрюса

| Критерии | | Тип программного обеспечения | | |
|--|------------------------------------|---|--|----------------|
| | | Библиотека для языка программирования | Десктопное приложения | Веб-приложение |
| Наличие в исследовательской команде опыта программирования | Опыт программирования отсутствует | | «NovoSpark Visualizer» (разработка NovoSpark Corporation) | |
| | Опыт программирования присутствует | <i>matplotlib (python), andrewsplot (MATLAB), andrews (R)</i> | «NovoSpark Visualizer» (разработка NovoSpark Corporation) | |

В результате построения карты сегментирования выявлено, какие ниши на рынке услуг не заняты конкурентами или где уровень конкуренции низок. Такими нишами являются веб-приложения для команд как с опытом, так и без опыта программирования. Поэтому было решено разрабатывать программную систему в виде веб-приложения, ориентированного на исследовательские команды без опыта программирования, что, однако, не отменяет возможности его использования командами с опытом программирования.

4.1.2 Анализ конкурентных технических решений

Необходимо систематически проводить анализ конкурирующих разработок, существующих на рынке. Это способствует своевременному усовершенствованию научного исследования и успешному противостоянию соперникам. Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения позволяет провести оценку эффективности разработки и определить направления для дальнейшего усовершенствования.

Для сравнительного анализа были выбрано несколько вариантов существующего программного обеспечения: библиотеки для языков программирования и десктопное приложение «*NovoSpark Visualizer*». Библиотеки для различных языков программирования имеют схожие конкурентные характеристики, поэтому для них приводятся общая оценка. Приложение «*NovoSpark Visualizer*» является наиболее близким по функциональным возможностям и характеристикам к разрабатываемой системе, поэтому рассматривается как потенциальное конкурентное решение. Оценочная карта представлена в таблице 4. В таблице B_{ϕ} и K_{ϕ} относятся к разрабатываемой системе, B_{k1} и K_{k1} к библиотекам для языков программирования, B_{k2} и K_{k2} к приложению «*NovoSpark Visualizer*».

Таблица 4 – Оценочная карта для сравнения конкурентных технических решений

| Критерии оценки | Вес критерия | Баллы | | | Конкурентоспособность | | |
|---|--------------|----------------|-----------------|-----------------|-----------------------|-----------------|-----------------|
| | | Б _ф | Б _{к1} | Б _{к2} | К _ф | К _{к1} | К _{к2} |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Технические критерии оценки ресурсоэффективности | | | | | | | |
| 1. Повышение производительности труда пользователя | 0.15 | 5 | 4 | 5 | 0.75 | 0.6 | 0.75 |
| 2. Удобство в эксплуатации | 0.20 | 5 | 2 | 4 | 1 | 0.4 | 0.8 |
| 3. Надежность | 0.15 | 5 | 4 | 4 | 0.75 | 0.6 | 0.6 |
| 4. Функциональные возможности | 0.15 | 3 | 5 | 5 | 0.45 | 0.75 | 0.75 |
| Экономические критерии оценки эффективности | | | | | | | |
| 1. Конкурентоспособность продукта | 0.15 | 4 | 4 | 3 | 0.4 | 0.4 | 0.3 |
| 2. Уровень востребованности среди потребителей | 0.20 | 4 | 4 | 4 | 0.4 | 0.4 | 0.4 |
| 3. Цена | 0.15 | 4 | 5 | 3 | 0.2 | 0.25 | 0.15 |
| 4. Послепродажное обслуживание | 0.15 | 4 | 5 | 0 | 0.2 | 0.25 | 0 |
| 5. Финансирование научной разработки | 0.15 | 0 | 0 | 0 | 0 | 0 | 0 |
| Итого | 1 | | | | 4.15 | 3.65 | 3,75 |

Анализ конкурентных технических решений определяется по следующей формуле:

$$K = \sum B_i * B_i, \quad (9)$$

где K – конкурентоспособность научной разработки или конкурента;

B_i – вес показателя (в долях единицы);

B_i – балл i -го показателя.

Исходя из расчетов, можно сделать вывод, что разрабатываемая система визуализации многомерных данных на основе кривых Эндрюса имеет

достаточно высокий уровень конкурентоспособности. Она имеет следующие преимущества: удобство в эксплуатации, надежность, а также повышение производительности труда. Позиции конкурентов наиболее уязвимы в надежности и удобстве использования, что определяет конкурентное преимущество разрабатываемой системы.

4.1.3 SWOT-анализ

В целях исследования внешней и внутренней среды объекта был проведен SWOT-анализ, который отражает сильные и слабые стороны, возможности и угрозы разрабатываемого продукта (таблица 5, таблица 6).

Таблица 5 – Матрица SWOT

| | | |
|---|---|---|
| | <p>Сильные стороны:</p> <p>С1 - Опыт работы со средствами разработки, необходимыми для реализации программной системы.</p> <p>С2 - Имеющийся опыт работы с данными, их визуализации.</p> <p>С3 - Имеющийся опыт создания пользовательских интерфейсов.</p> <p>С4 - Имеющийся опыт разработки веб-приложений.</p> | <p>Слабые стороны:</p> <p>Сл1 - Отсутствие опыта привлечения новых пользователей.</p> <p>Сл2 - Малый опыт в организации информационной безопасности программного обеспечения.</p> <p>Сл3 – Отсутствие финансирования.</p> <p>Сл4 - Отсутствие ресурсов для проведения покрывающего ручного тестирования.</p> |
| <p>Возможности:</p> <p>В1 - Отсутствие аналогичного программного обеспечения, предоставляющего возможность применения метода кривых Эндрюса.</p> <p>В2 - Возрастающий интерес к различным областям анализа данных.</p> <p>В3 - Рост числа исследовательских команд.</p> | <p>За счет отсутствия аналогов системы в сегменте веб-приложений она имеет высокую конкурентоспособность.</p> <p>Наличие опыта, необходимого для реализации системы, делает возможным ее реализацию в короткий срок, что может позволить выйти на рынок без конкурентов.</p> | <p>Отсутствие опыта привлечения новых пользователей может замедлить рост аудитории системы, что может нивелировать благоприятное состояние внешней среды проекта.</p> |

Продолжение таблицы 5

| | | |
|--|---|--|
| Угрозы: У1 - Появление конкурирующих программных систем. У2 - Уменьшение количества исследовательских команд. У3 - Появление метода кривых Эндрюса в существующих системах визуализации. У4 - Отсутствие интереса исследователей к методу кривых Эндрюса. | <p>Низкий уровень конкуренции, имеющийся на данный момент в предполагаемом сегменте рынка, со временем может увеличиться. Для успешного развития проекта необходимо соблюсти баланс между ранним выходом системы на рынок и качеством продукта.</p> | <p>Ограниченность ресурсов, связанная с отсутствием финансирования, может сказаться на времени реализации системы и ее конечном качестве. Для обеспечения высокой конкурентоспособности необходимо найти источники финансирования.</p> |
|--|---|--|

Таблица 6 – Интерактивная матрица проекта

| | | Сильные стороны проекта | | | | Слабые стороны проекта | | | |
|---------------------|----|-------------------------|----|----|----|------------------------|-----|-----|-----|
| | | С1 | С2 | С3 | С4 | Сл1 | Сл2 | Сл3 | Сл4 |
| Возможности проекта | B1 | + | + | + | + | + | - | - | - |
| | B2 | - | - | - | - | + | - | - | - |
| | B3 | - | - | - | - | - | - | - | - |
| Угрозы проекта | У1 | + | + | + | + | + | + | - | - |
| | У2 | - | - | - | - | - | - | - | - |
| | У3 | + | + | + | + | + | + | - | - |
| | У4 | - | - | + | - | + | - | - | - |

Таким образом, в результате *SWOT*-анализа были выявлены слабые и сильные стороны, а также возможные варианты повышения эффективности и минимизации угроз в проекте по разработке программной системы визуализации многомерных данных на основе кривых Эндрюса.

4.2 Планирование научно-исследовательских работ

4.2.1 Структура работ в рамках научного исследования

Планирование комплекса предполагаемых работ осуществляется в следующем порядке:

- определение структуры работ в рамках научного исследования;
- определение участников каждой работы;
- установление продолжительности работ;
- построение графика проведения научных исследований.

Для выполнения научного исследования формируется рабочая группа, по каждому виду запланированных работ устанавливается соответствующая должность исполнителей. Перечень этапов и работ, распределение исполнителей по видам работ приведен в таблице 7.

Таблица 7 – Перечень этапов, работ и распределение исполнителей

| Основные этапы | № работ | Содержание работ | Должность исполнителя |
|------------------------------------|---------|---|-----------------------|
| Разработка технического задания | 1 | Выбор темы научного исследования | Студент, руководитель |
| | 2 | Составление и утверждение технического задания | Студент, руководитель |
| | 3 | Календарное планирование работ по теме исследования | Руководитель |
| Анализ предметной области | 4 | Подбор и изучение материалов по теме | Студент |
| Выбор методов реализации проекта | 5 | Выбор технологий разработки | Студент, руководитель |
| | 6 | Выбор инструментов разработки | Студент, руководитель |
| Проектирование программной системы | 7 | Проектирование пользовательского интерфейса | Студент |
| | 8 | Проектирование архитектуры системы | Студент |

Продолжение таблицы 7

| Основные этапы | № работ | Содержание работ | Должность исполнителя |
|-----------------------------------|---------|--|-----------------------|
| Разработка программной системы | 9 | Разработка модуля преобразования данных в кривые Эндрюса | Студент |
| | 10 | Внедрение авторизации пользователей | Студент |
| Обобщение и оценка результатов | 11 | Оценка полученной системы | Студент, руководитель |
| Оформление отчета по исследованию | 12 | Составление пояснительной записки | Студент |
| | 13 | Подготовка презентационного материала | Студент |

4.2.2 Определение трудоемкости выполнения работ

Трудовые затраты в большинстве случаев образуют основную часть стоимости разработки, поэтому важным моментом является определение трудоемкости работ каждого из участников научного исследования. Трудоемкость выполнения научного исследования оценивается экспертным путем в человеко-днях и носит вероятностный характер, т.к. зависит от множества трудно учитываемых факторов.

Ожидаемое (среднее) значение трудоемкости $t_{ож\ i}$ вычисляется по формуле:

$$t_{ож\ i} = \frac{3t_{min\ i} + 2t_{max\ i}}{5}, \quad (10)$$

где $t_{ож\ i}$ – ожидаемая трудоемкость выполнения i -ой работы чел.-дн.;

$t_{min\ i}$ – минимально возможная трудоемкость выполнения заданной i -ой работы (оптимистическая оценка: в предположении наиболее благоприятного стечения обстоятельств), чел.-дн.;

$t_{max\ i}$ – максимально возможная трудоемкость выполнения заданной i -ой работы (пессимистическая оценка: в предположении наиболее неблагоприятного стечения обстоятельств), чел.-дн.

Исходя из ожидаемой трудоемкости работ, определяется продолжительность каждой работы в рабочих днях T_{pi} , учитывающая параллельность выполнения работ несколькими исполнителями.

$$T_{pi} = \frac{t_{ож\ i}}{Ч_i}, \quad (11)$$

где T_{pi} – продолжительность одной работы, раб. дн.;

$t_{ож\ i}$ – ожидаемая трудоемкость выполнения одной работы, чел.-дн.

$Ч_i$ – численность исполнителей, выполняющих одновременно одну и ту же работу на данном этапе, чел.

Расчеты ожидаемой трудоемкости и продолжительности работ представлены в таблице 8.

4.2.3 Разработка графика проведения научного исследования

Для удобства построения графика длительность каждого из этапов работ следует перевести из рабочих в календарные дни. Для этого необходимо воспользоваться следующей формулой:

$$T_{ki} = T_{pi} \times k_{кал}, \quad (12)$$

где T_{ki} – продолжительность выполнения i -й работы в календарных днях;

T_{pi} – продолжительность выполнения i -й работы в рабочих днях;

$k_{кал}$ – коэффициент календарности.

Коэффициент календарности определяется по следующей формуле:

$$k_{кал} = \frac{T_{кал}}{T_{кал} - T_{вых} - T_{пр}} = \frac{366}{248} = 1,4718, \quad (13)$$

где $T_{кал}$ – календарные дни;

$T_{вых}$ – выходные дни;

$T_{пр}$ – праздничные дни.

Тогда длительность каждого из этапов работ в календарных днях будет равна

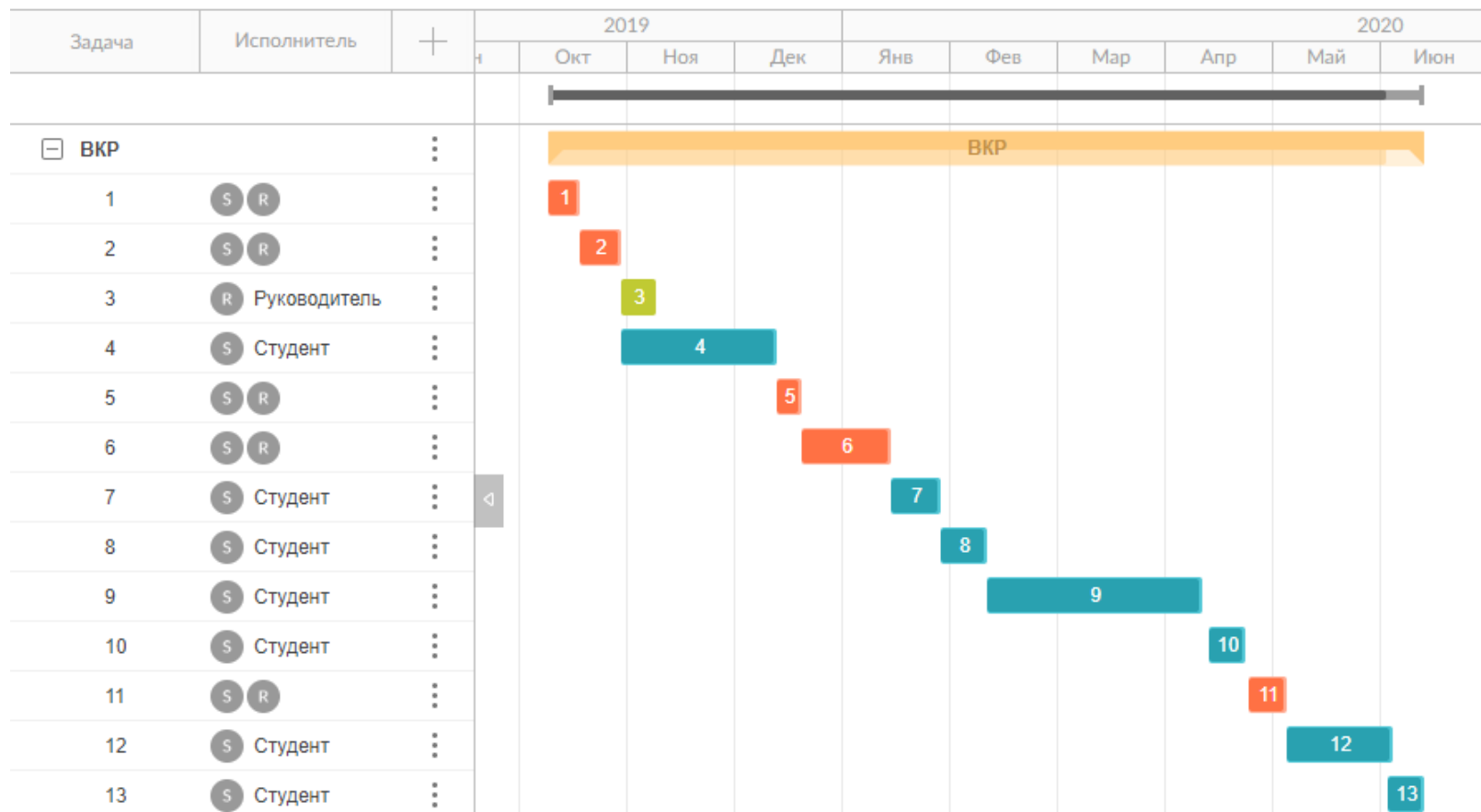
$$T_{ki} = T_{pi} \times 1,4718.$$

Временные показатели проведения научного исследования представлены в таблице 8.

Таблица 8 – Временные показатели проведения научного исследования

| № работ | Трудоемкость работ | | | Исполнители | Длительность работ в рабочих днях T_{pi} | Длительность работ в календарных днях T_{ki} |
|---------|-------------------------|-------------------------|---------------------------|--------------------------|---|---|
| | t_{min} , чел.-дн. | t_{max} , чел.-дн. | $t_{ож\ i}$, чел.-дн. | | | |
| 1 | 9,0 | 18,0 | 12,6 | Студент, руководитель | 6,3 | 18,6 |
| 2 | 9,0 | 22,5 | 14,4 | Студент, руководитель | 7,2 | 21,2 |
| 3 | 4,5 | 11,3 | 7,2 | Руководитель | 7,2 | 10,6 |
| 4 | 24,3 | 35,6 | 28,8 | Студент | 28,8 | 42,5 |
| 5 | 6,3 | 13,1 | 9 | Студент, руководитель | 4,5 | 13,3 |
| 6 | 25,2 | 43,3 | 32,5 | Студент, руководитель | 16,2 | 47,8 |
| 7 | 6,3 | 13,1 | 9 | Студент | 9 | 13,3 |
| 8 | 6,3 | 10,8 | 8,1 | Студент | 8,1 | 11,9 |
| 9 | 18,0 | 72,1 | 39,7 | Студент | 39,7 | 58,4 |
| 10 | 6,3 | 8,6 | 7,2 | Студент | 7,2 | 10,6 |
| 11 | 6,3 | 22,1 | 12,6 | Студент, руководитель | 6,3 | 18,6 |
| 12 | 16,2 | 25,2 | 19,8 | Студент | 19,8 | 29,2 |
| 13 | 6,3 | 8,6 | 7,2 | Студент | 7,2 | 10,6 |
| Итого | Всего | | | | 167,7 | 247 |
| | Руководитель | | | | 88,3 | 130 |
| | Студент | | | | 160,5 | 236,2 |

На основании таблицы 8 строится календарный план-график, представленный на рисунке 38 в виде диаграммы Ганта.



Руководитель
и студент



Руководитель



Студент

Рисунок 38 – Диаграмма Ганта

На рисунке 38 работы выделены цветом в зависимости от исполнителей. Некоторые работы могут выполняться одновременно.

В таблице 8, а также на рисунке 38, введены обозначения: студент – Ширыкалов Анатолий Максимович, руководитель – Марухина Ольга Владимировна.

4.3 Бюджет научно-технического исследования (НТИ)

При планировании бюджета НТИ должно быть обеспечено полное и достоверное отражение всех видов расходов, связанных с его выполнением. В процессе формирования бюджета НТИ используются следующие статьи затрат:

- материальные затраты НТИ;
- основная заработная плата исполнителей;
- дополнительная заработная плата исполнителей;
- отчисления во внебюджетные фонды (страховые отчисления);
- накладные расходы.

4.3.1 Расчет материальных затрат НТИ

Данная статья включает стоимость всех материалов, используемых при разработке проекта, например, сырьё и материалы, запасные части для ремонта оборудования, комплектующие, канцелярские принадлежности.

В ходе выполнения работы использовались канцелярские принадлежности, общей суммой 500 рублей. Эта сумма также включает в себя расходы на распечатку необходимых материалов для проведения работы и её проверки.

4.3.2 Расчет затрат на специальное оборудование для научных (экспериментальных) целей

В данную статью включают все затраты, связанные с приобретением специального оборудования (приборов, контрольно-измерительной аппаратуры, стендов, устройств и механизмов), необходимого для проведения работ по конкретной теме проекта.

Рассчитаем амортизацию использованного оборудования. Для выполнения работы использовался ПК с первоначальной стоимостью 60000 рублей; срок полезного использования для машин офисных код 330.28.23.23 составляет 2-3 года; ПК использовался для написания ВКР в течение 8 месяцев. Тогда:

- норма амортизации:

$$A_n = \frac{1}{n} * 100\% = \frac{1}{3} \times 100\% = 33,33\%$$

- годовые амортизационные отчисления:

$$A_g = 60000 \times 0,33 = 19800 \text{ рублей}$$

- ежемесячные амортизационные отчисления:

$$A_m = \frac{19800}{12} = 1650 \text{ рублей}$$

- итоговая сумма амортизации основных средств:

$$A = 1650 \times 8 = 13200 \text{ рублей}$$

Затраты на амортизацию приведены в таблице 9.

Таблица 9 – Расчет затрат на амортизацию

| Наименование | Затраты, руб. |
|----------------|---------------|
| Амортизация ПК | 13200 |

4.3.3 Основная заработная плата исполнителей темы

В данную статью включается основная заработная плата научных и инженерно-технических работников, рабочих макетных мастерских и опытных производств, непосредственно участвующих в выполнении работ по данной теме.

Для расчета основной заработной платы студента берем оклад, равный окладу ассистента без степени, т.е. 21760 руб. Оклад руководителя составляет 33664 (доцент, кандидат технических наук).

Заработная плата основная:

$$З_{осн} = З_{дн} \times T_p \times (1 + K_{пр} + K_d) \times K_p, \quad (14)$$

где: $З_{дн}$ – среднедневная заработная плата, руб.;

$K_{пр}$ – премиальный коэффициент (0,3);

K_d – коэффициент доплат и надбавок (0,2-0,5);

K_p – районный коэффициент (для Томска 1,3);

T_p – продолжительность работ, выполняемых работником, раб. дни.

Среднедневная заработная плата:

$$З_{дн} = \frac{З_m \times M}{F_d}, \quad (15)$$

где: $З_m$ – месячный оклад работника, руб.

M – количество месяцев работы без отпуска в течение года: при отпуске в 24 раб. дня $M = 11,2$ месяца, 5-дневная неделя;

F_d – действительный годовой фонд рабочего времени персонала, раб. дн. (табл. 10).

Таблица 10 – Баланс рабочего времени (для 6-дневной недели)

| Показатели рабочего времени | Дни |
|--|-----|
| Календарные дни | 366 |
| Нерабочие дни (праздники/выходные) | 118 |
| Потери рабочего времени (отпуск/невыходы по болезни) | 24 |
| Действительный годовой фонд рабочего времени | 224 |

Среднедневная заработная плата студента:

$$З_{\text{дн}} = \frac{З_{\text{м}} \times М}{F_{\text{д}}} = \frac{21760 \times 11,2}{224} = 1088 \text{ руб.}$$

Среднедневная заработная плата руководителя:

$$З_{\text{дн}} = \frac{З_{\text{м}} \times М}{F_{\text{д}}} = \frac{33664 \times 11,2}{224} = 1683,2 \text{ руб.}$$

С учетом всех повышающих коэффициентов основная заработная плата студента:

$$З_{\text{осн}} = 1088 * 160,5 * (1 + 0,3 + 0,3) * 1,3 = 363\,217,92 \text{ руб.}$$

С учетом всех повышающих коэффициентов основная заработная плата руководителя:

$$З_{\text{осн}} = 1683,2 * 88,3 * (1 + 0,3 + 0,3) * 1,3 = 309\,143,24 \text{ руб.}$$

Расчет основной заработной платы приведен в таблице 11.

Таблица 11 – Расчет основной заработной платы

| Исполнители | З _{дн} , руб. | К _{пр} | К _д | К _р | Т _р | З _{осн} , руб. |
|-----------------|------------------------|-----------------|----------------|----------------|----------------|-------------------------|
| Ширыкалов А. М. | 1088 | 0,3 | 0,3 | 1,3 | 160,5 | 363 217,92 |
| Марухина О. В. | 1683,2 | 0,3 | 0,3 | 1,3 | 88,3 | 309 143,24 |
| Итого: | | | | | | 672 361,16 |

4.3.4 Дополнительная заработная плата исполнителей темы

Затраты по дополнительной заработной плате исполнителей темы учитывают величину предусмотренных Трудовым кодексом РФ доплат за отклонение от нормальных условий труда, а также выплат, связанных с обеспечением гарантий и компенсаций (при исполнении государственных и

общественных обязанностей, при совмещении работы с обучением, при предоставлении ежегодного оплачиваемого отпуска и т.д.).

Расчет дополнительной заработной платы ведется по следующей формуле:

$$З_{\text{доп}} = k_{\text{доп}} \cdot З_{\text{осн}}, \quad (16)$$

где $k_{\text{доп}}$ – коэффициент дополнительной заработной платы (на стадии проектирования принимается равным 0,12 – 0,15).

Дополнительная заработная плата студента:

$$З_{\text{доп}} = 0,13 \cdot 363\,217,92 = 47\,218,32 \text{ руб.}$$

Дополнительная заработная плата руководителя:

$$З_{\text{доп}} = 0,13 \cdot 309\,143,24 = 40\,188,62 \text{ руб.}$$

Суммарная дополнительная заработная плата:

$$З_{\text{доп}} = 47\,218,32 + 40\,188,62 = 87\,406,94 \text{ руб.}$$

4.3.5 Отчисления во внебюджетные фонды (страховые отчисления)

В данной статье расходов отражаются обязательные отчисления по установленным законодательством Российской Федерации нормам органам государственного социального страхования (ФСС), пенсионного фонда (ПФ) и медицинского страхования (ФФОМС) от затрат на оплату труда работников.

Величина отчислений во внебюджетные фонды определяется исходя из формулы 17:

$$З_{\text{внеб}} = k_{\text{внеб}} \cdot (З_{\text{осн}} + З_{\text{доп}}), \quad (17)$$

где $k_{\text{внеб}}$ – коэффициент отчислений на уплату во внебюджетные фонды (пенсионный фонд, фонд обязательного медицинского страхования и пр.).

Таким образом, суммарная величина отчислений во внебюджетные фонды составляет:

$$\begin{aligned}
Z_{\text{внеб}} &= 0,3 * ((363\,217,92 + 47\,218,32) + (309\,143,24 + 40\,188,62)) \\
&= 227\,930,43 \text{ руб.}
\end{aligned}$$

4.3.6 Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: оплата услуг связи, электроэнергии, почтовые и телеграфные расходы, размножение материалов и т.д.

Величина данной статьи расходов определяется по формуле 18:

$$Z_{\text{накл}} = 0,16 * (Z_{\text{м}} + Z_{\text{осн}} + Z_{\text{доп}} + Z_{\text{внеб}} + Z_{\text{ам}}) \quad (18)$$

Таким образом, накладные расходы составляют:

$$\begin{aligned}
Z_{\text{накл}} &= 0,16 * (500 + 672\,361,16 + 87\,406,94 + 227\,930,43 + 13\,200) \\
&= 160\,223,76 \text{ руб.}
\end{aligned}$$

4.3.7 Формирование бюджета затрат научно-исследовательского проекта

Расчет бюджета затрат НТИ приведен в таблице 12.

Таблица 12 – Расчет бюджета затрат НТИ

| Наименование | Сумма, руб. | Удельный вес, % |
|--|--------------|-----------------|
| Материальные затраты | 500 | 0,04 |
| Затраты на специальное оборудование | 13 200 | 1,14 |
| Затраты на основную заработную плату | 672 361,16 | 57,88 |
| Затраты на дополнительную заработную плату | 87 406,94 | 7,52 |
| Страховые взносы | 227 930,43 | 19,62 |
| Накладные расходы | 160 223,76 | 13,79 |
| Общий бюджет | 1 161 622,29 | 100 |

Основной статьей расхода является основная заработная плата, т.к. составляет больше половины от всех затрат НТИ.

4.4 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Определение эффективности происходит на основе расчета интегрального показателя эффективности научного исследования. Его нахождение связано с определением двух средневзвешенных величин: финансовой эффективности и ресурсоэффективности.

Интегральный финансовый показатель разработки определяется как:

$$I_{\Phi}^p = \frac{\Phi_p}{\Phi_{max}}, \quad (19)$$

где: Φ_p – стоимость исполнения, руб.;

Φ_{max} – максимальная стоимость исполнения научно-исследовательского проекта (в т. ч. аналоги).

$$I_{\Phi}^p = \frac{1\,161\,622,29}{1\,161\,622,29} = 1$$

Полученная величина интегрального финансового показателя разработки отражает соответствующее численное удешевление стоимости разработки в размах.

Интегральный показатель ресурсоэффективности исполнения объекта исследования можно определить следующим образом:

$$I_p = \sum_n a \times b, \quad (20)$$

где a – весовой коэффициент параметра;

b – бальная оценка параметра для аналога и разработки, устанавливается экспертным путем по выбранной шкале оценивания;

n – число параметров сравнения.

Расчет интегрального показателя ресурсоэффективности приведен в таблице 13.

Таблица 13 – Расчет интегрального показателя ресурсоэффективности

| Критерии \ Объект исследования | Весовой коэффициент параметра | Оценка выполнения |
|---|-------------------------------|-------------------|
| 1. Способствует росту производительности труда пользователя | 0,25 | 4 |
| 2. Удобство в эксплуатации (соответствует требованиям потребителей) | 0,2 | 5 |
| 3. Функциональные возможности | 0,2 | 4 |
| 4. Надежность | 0,2 | 4 |
| 5. Экономия времени | 0,15 | 5 |
| Итого | 1 | |

$$I_p = 0,25 * 4 + 0,2 * 5 + 0,2 * 4 + 0,2 * 4 + 0,15 * 5 = 4,35$$

Интегральный показатель эффективности исполнения разработки ($I_{\text{исп}}$) определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{\text{исп}} = \frac{I_p}{I_{\text{ф}}} = \frac{4,35}{1} = 4.35. \quad (21)$$

Результат работы можно считать положительным, так как оценка интегрального показателя ресурсоэффективности выше 4.

Таким образом, полученные при анализе конкурентных решений данные позволяют сделать вывод, что разработка является привлекательной для инвесторов. Продукт имеет много преимуществ перед рассмотренными конкурентами, в особенности по таким критериям, как удобство в эксплуатации, надежность и повышение производительности труда. *SWOT*-анализ позволил выявить слабые и сильные стороны, возможные перспективы и угрозы, а также предложены рекомендации по минимизации их влияния.

Также была построена структура работ проекта и определены ответственные должности для их выполнения. В соответствии с назначенными работами была рассчитана их трудоемкость и составлен план-график работ в виде диаграммы Ганта. Общая длительность проектирования и разработки программного продукта составила 247 дней. Общий бюджет НТИ составил 1 161 622,29 рублей. Бюджет включает в себя затраты на основную и дополнительную заработную плату работников, материальные затраты, отчисления во внебюджетные фонды и накладные расходы.

5 Социальная ответственность

Визуализация является важной частью первичной обработки данных, так как позволяет исследователю произвести их качественный анализ. Существует множество различных методов визуализации, однако многие из них становятся малоэффективными при увеличении размерности данных. Одним из методов, позволяющих представить данные любой размерности на плоскости, является построение кривых Эндрюса.

Исследовательская работа заключалась в проектировании и разработке программной системы визуализации многомерных данных на основе метода кривых Эндрюса. Используя приложение, пользователи, не имеющие опыта программирования, могут получать визуализации наборов данных на основе метода кривых Эндрюса.

Согласно требованиям, предъявленным к разрабатываемой программной системе, для работы с приложением пользователями могут быть использованы стационарные персональные компьютеры и ноутбуки. Сама разработка при этом велась с использованием ноутбука.

5.1 Правовые и организационные вопросы обеспечения безопасности

5.1.1 Организационные мероприятия обеспечения безопасности

Разработка программной системы велась в помещении общей площадью 26 квадратных метров, в качестве искусственного источника освещения использовались 3 лампы накаливания, общей мощностью 85 Вт. Помещение оборудовано восемью компьютерными столами с выдвижной подставкой для клавиатуры, являющимися персональным рабочим местом. Автор научно-исследовательской работы взаимодействовал с электронно-вычислительными устройствами, находясь в операторском кресле, выполненном в виде компьютерного кресла с регулируемыми подлокотниками и углом наклона спинки. Перемещение кресла внутри помещения обеспечивают 5 пластиковых

колес диаметром 50 мм. Доступ к свободному креслу в помещении был постоянно.

С целью минимизации воздействия вредных факторов как при проектировании и разработке программной системы визуализации многомерных данных на основе метода кривых Эндрюса, так и во время ее эксплуатации, рабочее место должно быть организовано с учетом требований ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования» и СанПиН 2.2.2/2.4.1340-03 «Гигиенические требования к персональным электронно-вычислительным машинам и организации работы».

Ниже приведены наиболее важные для соблюдения фрагменты стандарта (используется оригинальная нумерация пунктов соответствие с ГОСТ 12.2.032-78 «ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования»):

1. Подвижность кресла относительно пола или другой поверхности, на которой оно установлено, может не ограничиваться. В случае необходимости обеспечения строго определенного положения человека-оператора по отношению к средствам отображения информации и органам управления, а также в случае, если трудовая деятельность человека-оператора сопряжена с силовыми и резкими движениями, кресло должно быть фиксировано. При этом, в зависимости от характера трудовой деятельности оператора, должна быть обеспечена возможность изменения положения кресла или сиденья в горизонтальной плоскости с фиксацией его в нужном положении. При необходимости подвижность кресла должна задаваться также вращением кресла на 180-360° вокруг вертикальной оси опорной конструкции кресла с фиксацией в нужном положении.
2. При работе двумя руками органы управления размещают с таким расчетом, чтобы не было перекрещивания рук.

3. Помещения, где размещаются рабочие места с ПЭВМ, должны быть оборудованы защитным заземлением (занулением) в соответствии с техническими требованиями по эксплуатации.
4. В помещениях, оборудованных ПЭВМ, проводится ежедневная влажная уборка и систематическое проветривание после каждого часа работы на ПЭВМ.
5. Рабочие столы следует размещать таким образом, чтобы видеодисплейные терминалы были ориентированы боковой стороной к световым проемам, чтобы естественный свет падал преимущественно слева.
6. В качестве источников света при искусственном освещении следует применять преимущественно люминесцентные лампы типа ЛБ и компактные люминесцентные лампы (КЛЛ). При устройстве отраженного освещения в производственных и административно-общественных помещениях допускается применение металлогалогенных ламп. В светильниках местного освещения допускается применение ламп накаливания, в том числе галогенные.
7. Экран видеомонитора должен находиться от глаз пользователя на расстоянии 600-700 мм, но не ближе 500 мм с учетом размеров алфавитно-цифровых знаков и символов.

Автором научно-исследовательской работы были соблюдены в допустимой мере все требования, предусматриваемые государственным стандартом 12.2.032-78. Во время выполнения выпускной квалификационной работы не происходило случаев, несущих в себе угрозы для здоровья и жизни разработчика, а также представляющих опасность для окружающей среды.

5.2 Производственная безопасность

В данном пункте производится анализ вредных и опасных факторов, которые могут возникнуть на одном из этапов проектирования и разработки

программной системы визуализации многомерных данных на основе метода кривых Эндрюса.

Рассмотрены следующие факторы: отклонение показателей микроклимата и естественного света, недостаточная освещенность и повышенный уровень электромагнитного излучения. Таблица 14 содержит информацию о воздействии факторов в зависимости от этапов научного исследования.

Таблица 14 – Возможные опасные и вредные факторы

| Факторы (по ГОСТ 12.0.003-74) | Этапы ВКР | | | Использование ПС | Нормативные документы |
|---|----------------|------------|-------------------------|------------------|--|
| | Проектирование | Разработка | Формирование отчетности | | |
| Отклонение показателей микроклимата | + | + | + | + | СанПиН 2.2.4.548-96 |
| Отсутствие или недостаток естественного света | + | + | + | + | СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95* |
| Недостаточная освещенность рабочей зоны | + | + | + | + | СанПиН 2.2.1/2.1.1.1278–03 |
| Повышенный уровень электромагнитных излучений | + | + | + | + | СанПиН 2.2.4.3359-16 |

Как видно из таблицы 1, опасные и вредные факторы воздействовали на разработчика на всех этапах выполнения работы, а также могут воздействовать на пользователя во время использования разработанной системы визуализации многомерных данных на основе метода кривых Эндрюса.

5.2.1 Анализ опасных и вредных производственных факторов

В таблице 15 приведены параметры перечисленных в таблице 14 вредных факторов.

Таблица 15 – Влияние опасных и вредных факторов

| Фактор | Источник | Воздействие | Допустимые нормы |
|---|--|--|--|
| Отклонение показателей микроклимата | Отсутствие кондиционеров | Вялость, усталость, сниженная концентрация | Таблица 16 |
| Отсутствие или недостаток естественного света | Периодическая необходимость работы за ЭВМ в ночное время | Ухудшение зрения, усталость глаз | КЕО не ниже 1,2%-1,5% |
| Недостаточная освещенность рабочей зоны | Недостаточная мощность осветительных приборов | Ухудшение зрения, усталость глаз | Освещенность на рабочей поверхности от системы общего искусственного освещения 200-300 лк. |
| Повышенный уровень электромагнитных излучений | Компоненты персональных компьютеров и ноутбуков | Возможно возникновение рака | Напряженность электростатического поля не более 20 кВ/м |

Норма микроклимата является плавающим параметром и зависит от температуры помещения, поверхностей, влажности и скорости воздуха. Допустимые величины показателей микроклимата продемонстрированы в таблице 16.

Таблица 16 – Допустимые величины показателей микроклимата

| Период года | Температура воздуха, °С | Температура поверхностей, °С | Относительная влажность воздуха, % | Скорость движения воздуха, м/с |
|-------------|-------------------------|------------------------------|------------------------------------|--------------------------------|
| Холодный | 20,0-21,9 | 19,0-26,0 | 15-75 | 0,1 |
| Теплый | 21,0-22,9 | 20,0-29,0 | | |

5.2.2 Обоснование мероприятий по снижению уровней воздействия опасных и вредных факторов

5.2.2.1 Мероприятия по снижению воздействия недопустимого микроклимата

Для восстановления и поддержания допустимого микроклимата необходимо придерживаться следующих правил:

- Оборудование помещения системами обогрева, вентилирования и увлажнения.
- Оборудование помещения современными пластиковыми окнами, поддерживающими возможность микровентилирования.
- Защита фасада здания от солнца: шторы, жалюзи, навесы и т.д.
- Рационально размещать рабочие места.
- Ежедневная влажная уборка рабочего помещения.

5.2.2.2 Мероприятия по снижению воздействия недостатка освещенности рабочего места и естественного света

Для решения проблемы отсутствия или недостатка естественного света и плохой освещенности рабочего места подходят следующие пункты:

- Сокращение времени работы.
- Своевременная чистка стекол в светопроемах.
- Снос деревьев, препятствующих проникновению света в помещение.
- Ремонт помещения в светлых тонах.
- Установка более мощных ламп или в большем количестве.
- Установка ламп в правильном положении.

5.2.2.3 Мероприятия по снижению воздействия недопустимого уровня электромагнитного излучения

Повышенный уровень электромагнитных излучений можно избежать, если следовать следующим пунктам:

- Прекратить использование мониторов с электронно-лучевой трубкой.
- Использовать высокоэффективные блоки питания и прочие преобразователи напряжения.
- Располагать монитор в углу помещения для того, чтобы стены поглощали излучение.
- Выключать компьютер при его неиспользовании.
- Сокращать время, проводимое за компьютером.

5.3 Экологическая безопасность

Экологической безопасностью называется комплекс мероприятий по снижению негативных влияний производственной деятельности человека на окружающую среду и защиту человека от последствий этого влияния.

Для выполнения научно-исследовательской работы по проектированию и разработке программной системы визуализации многомерных данных на основе метода кривых Эндрюса, а также при эксплуатации разработанной системы использовались компьютеры и ноутбуки средней мощности. Современные электронно-вычислительные устройства не выбрасывают в окружающую среду каких-либо вредных веществ, однако используют для работы электроэнергию и создают электромагнитные поля. Производство и утилизация современных вычислительных устройств составляют серьезную проблему: текстолит, используемый при производстве микросхем, имеющих срок разложения более тысячи лет.

При производстве персональных компьютеров используются тяжелые, щелочноземельные металлы, ртуть, пластик и стекло, что без должной утилизации по окончании службы попадает в природу и остается в не переработанном виде.

Мероприятия, позволяющие сохранять экологическую безопасность находясь на своем рабочем месте:

- Правильная утилизация персональных компьютеров и ноутбуков, а также их комплектующих;
- Использование энергосберегающих ламп;
- Использование аккумуляторов вместо солевых батареек.

5.4 Безопасность в чрезвычайных ситуациях

Чрезвычайная ситуация (ЧС) – обстановка на определенной территории, сложившаяся в результате аварии, опасного природного явления, катастрофы, стихийного или иного бедствия, которая может повлечь за собой человеческие жертвы, а также ущерб здоровью человека или окружающей среде, значительные материальные потери и нарушение условий жизнедеятельности.

5.4.1 Пожар

Научно-исследовательская работа проходила в помещении, подходящем под определение офис. Одной из наиболее возможных чрезвычайных ситуаций, которая может возникнуть при работе в помещениях такого типа – пожар. К пожару могут привести неисправности в технических средствах, оргтехнике, а также действия самих сотрудников. Главное во время пожара – не поддаваться панике и действовать согласно правилам поведения при пожаре. Для сотрудника существует порядок действий и правила поведения в подобной чрезвычайной ситуации:

1. Заметив пожар или загорание, необходимо немедленно организовать оповещение об этом всех находящихся в здании людей, независимо от размеров и места пожара или загорания, равно как и при обнаружении хотя бы малейших признаков горения (дыма, запаха гари) и немедленно вызвать пожарную охрану по телефону «01». Очевидно, что быстрота прибытия пожарной помощи, позволит успешнее ликвидировать пожар и быстрее помочь людям, находящимся в опасности.
2. Сообщения о пожаре, как правило, передаются по телефону. Поэтому каждый человек должен хорошо знать места расположения телефонных аппаратов, особенно тех, которые доступны в любое время суток. Следует помнить, что с помощью сотового телефона можно вызвать помощь даже при отсутствии денег на счете или SIM-карты по номеру «112».

5.4.2 Землетрясение

Под землетрясением понимают подземные толчки и колебания земной поверхности. Землетрясения отражают процесс геологического преобразования планеты, первопричиной землетрясений являются глобальные геологические и тектонические силы

Томская область располагается на значительном удалении от зон сейсмической активности, однако, за последние 10 лет в пределах городской зоны произошло несколько землетрясений, максимальная зафиксированная амплитуда толчков составила 5,9. В связи с тем, что научно-исследовательская работа происходила в помещении на 6 этаже, что создает дополнительные факторы риска при земных толчках, участникам работ следует знать инструкции по поведению во время землетрясения:

1. При возможности захватить с собой документы, деньги, предметы первой необходимости, фонарик.
2. Остерегаться падающих предметов, оборванных проводов и других источников опасности.

3. Сохранять спокойствие и не допускать паники.
4. При нахождении на верхних этажах многоэтажного здания — оставаться в здании, предварительно открыть входную дверь, которая в дальнейшем может оказаться перекошенной и заклиненной.
5. Быстро занять наиболее безопасное место в помещении: в дверных проемах капитальных стен, у ближайшей к центру здания капитальной стены, опорной колонны, в углу комнаты, всегда подальше от окон, тяжелых предметов и мебели, которые могут опрокинуться.
6. В случае разрушения здания, сопровождающегося падением отдельных элементов перекрытия или частей капитальных стен, необходимо немедленно покинуть здание.
7. Покидая здание, не выпрыгивать из окон, расположенных выше первого этажа, стекла выбивать подручными средствами (стулом, табуреткой), в крайнем случае, рукой, обмотанной тряпкой.

Выводы по разделу

В результате изучения и анализа стандартов и правил, касающихся работы в помещениях с электронно-вычислительными устройствами, можно сделать вывод, что выполнение работы соответствовало всем заявленным нормам безопасности жизнедеятельности. Исполнитель не подвергался серьезному воздействию опасных факторов.

Рабочее место и помещение в целом во время проведения исследовательской работы соответствовало региональным стандартам, санитарно-эпидемиологическим правилам и нормам. Также приведенные правила, нормы и стандарты применимы к организации рабочего процесса при использовании разработанной программной системы визуализации многомерных данных на основе метода кривых Эндрюса.

Заключение

Несмотря на то, что разработка программной системы не была завершена, полученный результат уже имеет высокую ценность, так как позволяет оценить возможности метода кривых Эндрюса, а также метода кривых Эндрюса в совокупности с локальными вероятностями выброса.

В ходе работы был проведен анализ предметной области – работы с данными, а также изучен математический аппарат описанных методов.

Применимость метода кривых Эндрюса была исследована на примере набора данных «*Wheat Seeds Dataset*» [3], по результатам исследования был выполнен доклад на VI Международной научной конференции "Информационные технологии в науке, управлении, социальной сфере и медицине" 14 - 19 октября 2019г., в г. Томске, получивший диплом II степени, в сборнике материалов конференции были опубликованы тезисы доклада [7].

По промежуточным результатам разработки программной системы был выполнен доклад на XIII Всероссийской научной конференции молодых ученых "Наука. Технологии. Инновации." на секции "Математическое моделирование, анализ и обработка данных", 2-6 декабря 2019 г. в городе Новосибирске, с публикацией тезисов доклада [8] в сборнике конференции.

Была проведена работа по проектированию конечной версии программной системы, описаны модель данных системы и их потоки. Спроектированы бизнес-процессы работы с системой.

В ходе работы были разработаны основные функциональные модули программной системы: модуль работы с визуализациями, включающий модуль, отвечающий за применение метода локальной вероятности выброса, и модуль авторизации.

Перечисленные выше модули разработаны в соответствии со стандартом *OAuth2*, что гарантирует высокий уровень защищенности данных системы и пользователей.

По результатам проектирования и разработки модуля системы, отвечающего за применение метода локальной вероятности выброса, были составлены тезисы доклада на английском языке для участия в конференции XIII Международной Конференции «Новые Информационные Технологии в Исследовании Сложных Структур» *ISAM 2020*, принятые к публикации, однако в связи со сложной эпидемиологической ситуацией, проведение конференции было перенесено на сентябрь 2020 года.

Продолжение работы над программной системой, а именно:

- а) Разработка портала для управления данными в системе,
- б) Развертка системы для предоставления группе исследователей под руководством Марухиной Ольги Владимировны возможности работы с системой,

планируется в магистратуре.

Список литературы

1. Andrews, D. F. *Plots of High-Dimensional Data* // *Biometrics*. – 1972. – Т. 28. – № 1 – С. 125-136.
2. Грошев С. В., Пивоварова Н. В. Использование кривых Эндрюса для визуализации многомерных данных в задачах многокритериальной оптимизации // *Машиностроение и компьютерные технологии*. – 2015. – № 12 – С. 197-214.
3. M. Catatonics, J. Niewczas, P. Kulczycki, P.A. Kowalski, S. Lukasik, S. Zak A *Complete Gradient Clustering Algorithm for Features Analysis of X-ray Images* // *Information Technologies in Biomedicine*. – 2010. – С. 15-24.
4. Kriegel H.P., Kröger P., Schubert E., Zimek A. *LoOP: Local outlier probabilities* // *International Conference on Information and Knowledge Management, Proceedings*. – 2009. – С. 1649-1652.
5. *JavaScript.ru* | Введение в *JavaScript*. [Электронный ресурс]. // 2007-2020 Илья Кантор. URL: <https://learn.javascript.ru/intro> (Дата обращения 30.02.2020).
6. *FastAPI.tiangolo.com* | *FastAPI* [Электронный ресурс]. // 2019-2020 URL: <https://fastapi.tiangolo.com/> (Дата обращения 30.12.2019).
7. *Reactjs.org* | *Getting Started* [Электронный ресурс]. // 2013–2020. URL: <https://reactjs.org/docs/getting-started.html> (Дата обращения 24.03.2020).
8. *Sitepoint.com* | *5 React Architecture Best Practices* [Электронный ресурс]. // 2000–2020. URL: <https://www.sitepoint.com/react-architecture-best-practices/> (Дата обращения 24.03.2020).
9. Ширыкалов А. М. Визуализация многомерных данных с использованием кривых Эндрюса / А. М. Ширыкалов // *Информационные технологии в науке, управлении, социальной сфере и медицине : сборник научных трудов VI Международной конференции, 14-19 октября 2019 г., Томск. — Томск : Изд-во ТПУ, 2019. — [64-68].*
10. Разработка программной системы визуализации многомерных данных на

- основе кривых Эндрюса / А. М. Ширыкалов, О. В. Марухина // Наука. Технологии. Инновации сборник научных трудов XIII Всероссийской научной конференции молодых ученых , Новосибирск, 2-6 декабря 2019 г.: в 9 ч.: / Новосибирский государственный технический университет (НГТУ) . — 2019 . — Ч. 2 : Информационные технологии математического моделирования и обработки данных . — [С. 82-86]
11. *Hardt, D., Ed., "The OAuth 2.0 Authorization Framework", RFC 6749, DOI 10.17487/RFC6749, 2012, URL: <https://www.rfc-editor.org/info/rfc6749>* (Дата обращения 24.03.2020).
 12. Выброс [Электронный ресурс]. *URL: <https://ru.wikipedia.org/wiki/Выброс>* (Дата обращения 24.03.2020).
 13. *MATLAB* – Википедия [Электронный ресурс]. *URL: <https://ru.wikipedia.org/wiki/MATLAB>* (Дата обращения 24.03.2020).
 14. *Welcome to Python.org* [Электронный ресурс]. // 2001-2020. *URL: <https://python.org>* (Дата обращения 24.10.2019).
 15. ГОСТ 12.0.003-2015. Опасные и вредные производственные факторы. Классификация.
 16. Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 24.04.2020)
 17. СанПиН 2.2.2/2.4.1340-03 Гигиенические требования к персональным электронно-вычислительным машинам и организации работы.
 18. ТОИ Р-45-084-01. Типовая инструкция по охране труда при работе на персональном компьютере.
 19. ГОСТ 12.2.032-78 Система стандартов безопасности труда (ССБТ). Рабочее место при выполнении работ сидя. Общие эргономические требования.
 20. СанПиН 2.2.4.548-96 Гигиенические требования к микроклимату производственных помещений.
 21. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95.
 22. СанПиН 2.2.1/2.1.1.1278-03 Гигиенические требования к естественному,

искусственному и совмещенному освещению жилых и общественных зданий.

23. СанПиН 2.2.4.3359-16 Санитарно-эпидемиологические требования к физическим факторам на рабочих местах.

24. СанПиН 2.2.4.3359-16 Санитарно-эпидемиологические требования к физическим факторам на рабочих местах.