

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»**

Кафедра математического и компьютерного моделирования

Анализ экспериментальных данных

с помощью NoSQL

КУРСОВАЯ РАБОТА

студента 1 курса 147 группы
направления 09.03.03 — Прикладная информатика

механико-математического факультета

Иванова Дмитрия Алексеевича

Научный руководитель
доцент, к.т.н

И. А. Панкратов

Зав. кафедрой
зав. каф., д.ф.-м.н., доцент

Ю. А. Блинков

Саратов 2020

СОДЕРЖАНИЕ

Стр.

ВВЕДЕНИЕ	3
1 Краткая информация об OpenFOAM и ParaView	4
1.1 OpenFOAM	4
1.2 ParaView	6
2 Обзор существующих решений	8
2.1 HELYX OS	8
2.2 ANSA	9
2.3 CastNet	11
2.4 Итог	12
3 Проектирование информационной системы	13
3.1 Постановка задачи	13
3.2 Диаграмма прецедентов	13
3.3 Диаграмма классов	14
3.3.1 Диаграмма, представляющая способ организации экс- периментальных данных	14
3.3.2 Диаграмма для работы с базой данных и чтения экс- периментальных данных	16
3.3.3 Диаграмма классов для выполнения операций над гра- фиками	17
3.4 Диаграмма последовательностей	19
4 Выбор средств разработки	22
ЗАКЛЮЧЕНИЕ	23
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	24

ВВЕДЕНИЕ

При обработке экспериментальных данных, полученных в результате тематического моделирования физических процессов в CAD/CAE системах, особенно, когда проводится, например, серия экспериментов, в которых входные данные незначительно изменяются часто порождается большой объем результатов, подлежащих анализу или иными словами – пост-обработке. Причем, зачастую для анализа с полученными мало различающимися данным необходимо провести однотипные манипуляции. Учитывая все вышесказанное, становится ясна необходимость автоматизации такого процесса пост-обработки данных.

В курсовой работе будет спроектирована информационная система позволяющая упростить процесс анализа полученных экспериментальных данных. Для автоматизации будет использован язык программирования Python, для хранения результатов – NoSQL подход, а конкретно СУБД MongoDB.

Таким образом целью данной курсовой работы является проектирование информационной системы, автоматизирующей рутинные операции анализа экспериментальных данных. Задачи:

- Кратко рассмотреть CAD/CAE систему – OpenFOAM и ParaView.
- Рассмотреть существующие решения по данной тематике.
- Спроектировать UML-диаграммы для описания информационной системы.

1 Краткая информация об OpenFOAM и ParaView

1.1 OpenFOAM

OpenFOAM (англ. Open Source Field Operation And Manipulation CFD ToolBox) — открытая интегрируемая платформа для численного моделирования задач механики сплошных сред. [1]

Это пакет программ распространяемых свободно под лицензией GNU GPL, позволяющей решать задачи механики сплошных сред, в частности:

- Прочностные расчеты;
- Гидродинамика ньютоновских и неньютоновских вязких жидкостей как в несжимаемом,
- так и сжимаемом приближении с учётом конвективного теплообмена и действием сил гравитации. Для моделирования турбулентных течений возможно использование RANS-моделей, LES- и DNS-методов. Возможно решение дозвуковых, околосзвуковых и сверхзвуковых задач; Задачи теплопроводности в твёрдом теле;
- Многофазные задачи, в том числе с описанием химических реакций компонент потока;
- Задачи, связанные с деформацией расчётной сетки;
- Сопряжённые задачи;
- Некоторые другие задачи, при математической постановке которых требуется решение дифференциальных уравнений в частных производных в условиях сложной геометрии среды;

В основе кода лежит набор библиотек, предоставляющих инструменты для решения систем дифференциальных уравнений в частных производных как в пространстве, так и во времени. Рабочим языком кода является C++. OpenFOAM состоит из приблизительно 250 программ основанных на более чем 100 библиотеках. Каждое приложения выполняет свою конкретную задачу в рамках процесса расчета. Этапы работы представленные в соответствии с рисунком 1.1.

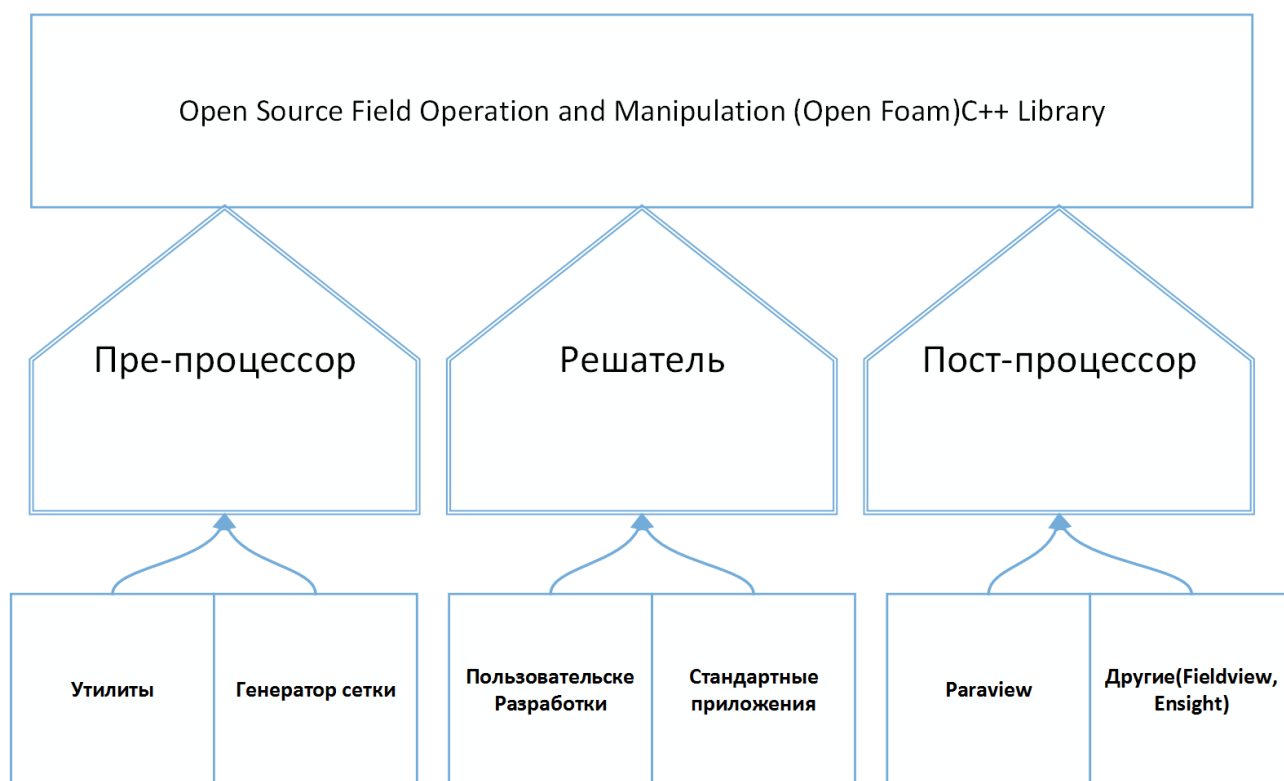


Рисунок 1.1 — Утилиты и программы входящие в пакет OpenFOAM, сгруппированные по этапам работы с расчетом

Работа с программой делится на три этапа:

1. Пре-процессинг;
2. Решение;
3. Пост-процессинг.

На этапе пре-процессинга в специальных файлах задаются входные данные для расчета примера, такие как: начальное время, конечное время, шаг и так далее. Также параметры для хранения решения: время, формат, тип сжатия. Также в препроцессинг включены настройки выбора различных схем расчета, которые влияют на точность и стабильность решения. После этого отдельно генерируется расчетная область (сетка), которая впоследствии может быть обработана различными утилитами [2]. Затем запускается решатель, который производит расчет. На этапе пост-процессинга полученные данные представляются в виде графиков. Также используются некоторые утилиты, например для конвертации из внутреннего формата OpenFOAM в широко используемый формат `vtk`.

1.2 ParaView

ParaView – открытый графический кросс-платформенный пакет для интерактивной визуализации в исследовательских целях, разрабатываемый Национальной Лабораторией Сандиа, компанией Kitware и Национальной Лабораторией Лос-Аламоса [3].

Пакет ParaView предоставляет пользователю возможности интерактивной визуализации и исследования больших массивов данных для качественного и количественного анализа.

Пакет может быть использован на компьютерах с операционными системами Windows, Linux, Mac OS X.

При разработке авторы придерживаются следующих целей:

- Открытость, кросс-платформенность — в пакете используются только открытые, мульти-платформенные технологии для визуализации данных.
- Поддержка различных, в том числе, гетерогенных вычислительных систем.
- Создание гибкого, интуитивного пользовательского интерфейса.

Таким образом, пакет ParaView во многом является скорее технологией обработки, чем всего лишь программным средством [4].

Некоторые возможности пакета:

- Визуализация расчетных областей.
- Визуализация полей (давление, скорость, температура, смещения и прочее).
- Построение срезов областей как плоскостью, так и заданной функцией.
- Построение изо-поверхностей.
- Построение векторных полей и линий тока.
- Позволяет показывать динамику развития протекающего процесса, отображая анимацию.

Основной формат данных ParaView – VTK, но пакет также содержит драйверы для работы с форматом OpenFOAM и поставляется вместе с дистрибутивом пакета.

Работа с ParaView может осуществляться как в интерактивном, так и пакетном режиме.

ParaView также предлагает богатый и мощный программный интерфейс на языке Python. Это позволяет пользователям автоматизировать обработку своих данных и использовать возможности, так называемого, набора инструментов визуализации – Visualization Tool Kit (VTK) [5].

2 Обзор существующих решений

Рассматриваемая в данной курсовой работе информационная система должна выполнять пост-обработку данных, полученных в результате численно эксперимента в пакете OpenFOAM, делая упор на автоматизацию функций для работы с серией данных. Рассмотрим доступные приложения осуществляющие автоматизацию рутинных функций в рамках пакета OpenFOAM.

2.1 HELYX OS

HELYX-OS - это графический пользовательский интерфейс с открытым исходным кодом, разработанный компанией ENGYS для работы со стандартными библиотеками OpenFOAM, предоставляемыми OpenFOAM Foundation и ESI-OpenCFD. Приложение предназначено для академического использования и работы с CFD начального уровня. Распространяется в соответствии с GNU General Public License [6].

HELYX-OS предоставляет полностью интерактивную, простую в использовании среду для выполнения всех задач предварительной обработки в процессе CFD, включая создание сетки, определение случая и выполнение решателя.

Существует также версия для корпоративного использования – CFD HELYX.

Преимущества:

- Встроенная поддержка как OpenFOAM, так и OpenFOAM+: возможность загружать существующие примеры, читая настройки непосредственно из доступных текстовых файлов проекта.
- Программа доступна на платформах Linux и Windows. Однако версия для Windows платна.
- Управление утилитой построения сеток snappyHexMesh, включая такие возможности как отображение геометрии и непосредственное построение прямо в окне приложения.
- Отдельный мониторинг решателя с отслеживанием остатков решения.

В корпоративной версии также следует выделить:

- Высокая масштабируемость.

- Возможность работы с использованием облачных технологий.
 - Модульность. Возможно расширение в рамках HELYX ADD-ONS,
- В соответствии с рисунком 2.1 изображен рабочий экран программы.

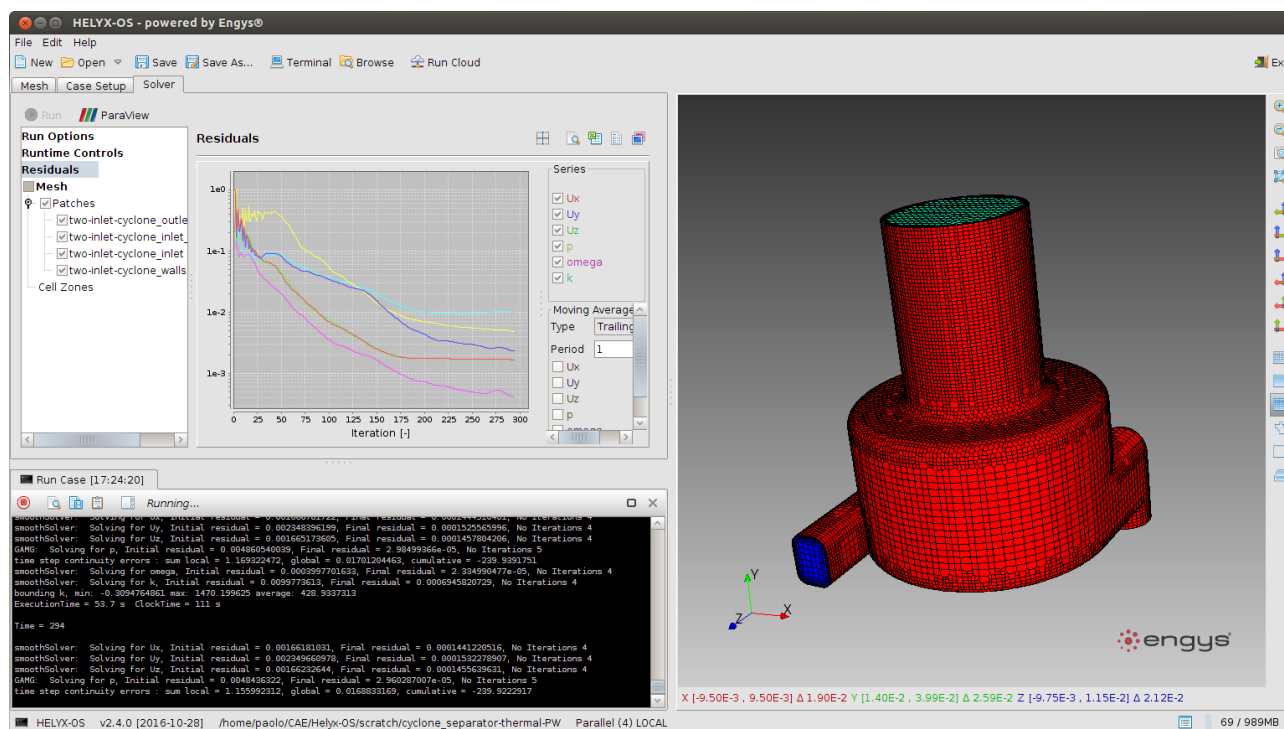


Рисунок 2.1 — Рабочий экран приложения Helyx OS

2.2 ANSA

ANSA - это инструмент пре-процессинга CAE, который предоставляет все необходимые функциональные возможности для построения полной модели, от CAD-данных до готового к вводу файла решателя, в единой интегрированной среде [7].

Все функции программного обеспечения размещены в интегрированной среде с настраиваемым графическим интерфейсом. Программное обеспечение доступно для всех современных популярных операционных систем в 32-битной и 64-битной архитектуре с использованием многоядерных процессоров.

Преимущества:

- Эффективная обработка данных для сложных структур моделей.

- Быстрое и качественное моделирование сложных геометрических моделей.
- Возможность взаимодействия между моделями, созданными для разных решателей.
- Высокоавтоматизированные процессы и инструменты настройки модели в одной программе.
- Уменьшены зависящие от пользователя подверженные ошибкам операции.
- Полное построение модели для многочисленных решателей в одной среде.

Рабочий экран приложения представлен в соответствии с рисунком 2.2.

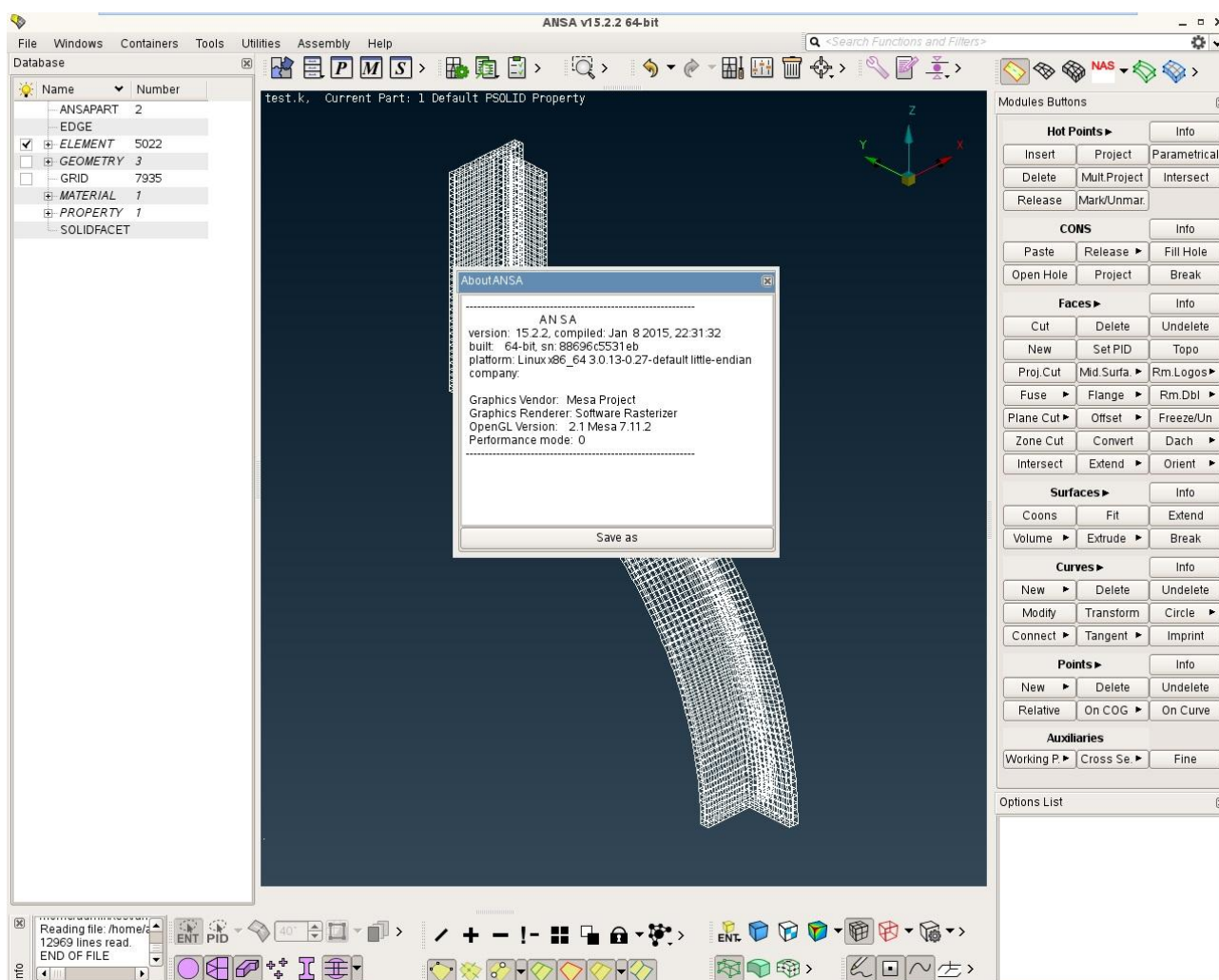


Рисунок 2.2 — Рабочий экран приложения ANSA

2.3 CastNet

CastNet упрощает использование технологических решений САЕ для решателей с открытым исходным кодом: кроме типичного редактирования текстовых файлов, предоставляется альтернативный способ работы с OpenFOAM на основе графического интерфейса, сохраняя полную совместимость со стандартными выпусками OpenFOAM. В результате рабочий процесс становится достаточно гибким, и пользователь может в любой момент переключаться между настройкой рабочего примера на основе текстового файла и графического интерфейса пользователя.

Вид рабочего экрана приложения представлен в соответствии с рисунком 2.3

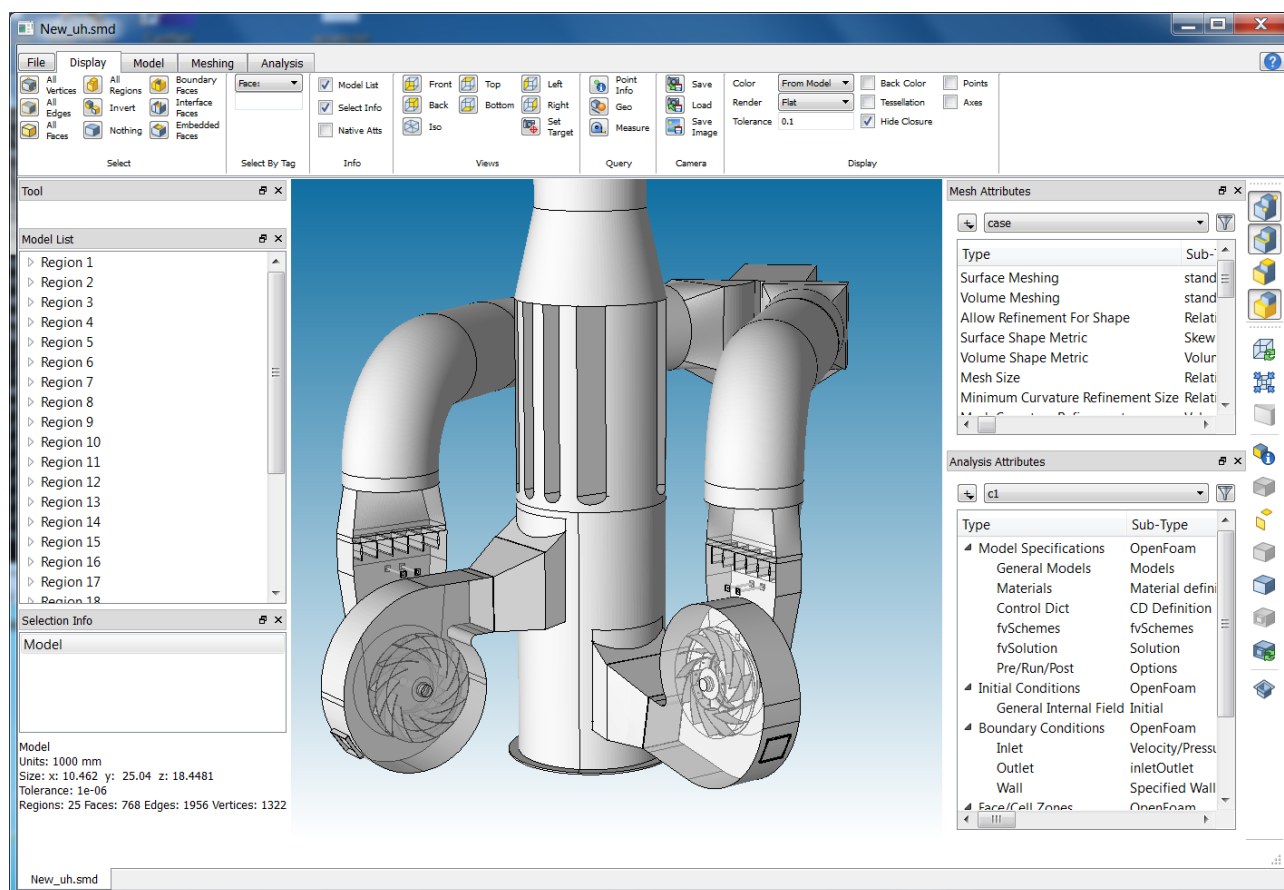


Рисунок 2.3 — Рабочий экран приложения CastNet

Ключевые особенности CastNet:

- Среда разработки на основе графического интерфейса пользователя, включающая предварительную обработку (создание сетки, настройку

примера), мониторинг решения и последующую обработку. Таким образом: доступ к мощным функциям решателя с открытым исходным кодом без редактирования текстовых файлов или необходимости детального изучения структуры ключевых слов OpenFOAM.

- Кроссплатформенное использование: поддержка среды для пакета программ OpenFOAM в операционных системах Windows и Linux.
- Библиотека шаблонов позволяет настраивать пример для более чем 30 решателей.
- Больше надежности в отношении результатов моделирования благодаря контролю сходимости.

2.4 Итог

Таким образом, рассмотренные существующие решения предлагают разнообразные и гибкие возможности по работе с примерами, однако ни одно из них не предоставляет функциональности по работе сразу с группой примеров, что в свою очередь вызывает трудности при анализе экспериментальных данных, которые состоят из набора примеров.

3 Проектирование информационной системы

3.1 Постановка задачи

Необходимо спроектировать приложение, которое бы выполняло процесс пост-обработки, то есть строило графики используя экспериментальные данные полученные из пакета программ OpenFOAM. Приложение должно также работать с группами экспериментальных данных, то есть выполнять конкретное действие построения графика, например срез, с группой из разных мало отличающихся примеров. Программа должна хранить экспериментальные данные и историю операций примера, также должна быть реализована возможность экспорта графиков в файлы. Для более подробного понимания информационной системы были построены UML-диаграммы.

3.2 Диаграмма прецедентов

Прецеденты – это технология определения функциональных требований к системе [9]. Диаграмма прецедентов (use case diagram) предназначена для описания взаимодействия проектируемой системы с любыми внешними или внутренними объектами - пользователями, другими системами и тому подобное. Основными понятиями при работе с диаграммой вариантов использования являются Актор (Actor) – это роль, которую выполняет пользователь или другая система, при взаимодействии с проектируемой системой. Вариант использования – это конечная единица взаимодействия актора и системы.

Диаграмма вариантов использования представлена в соответствии с рисунком 3.1

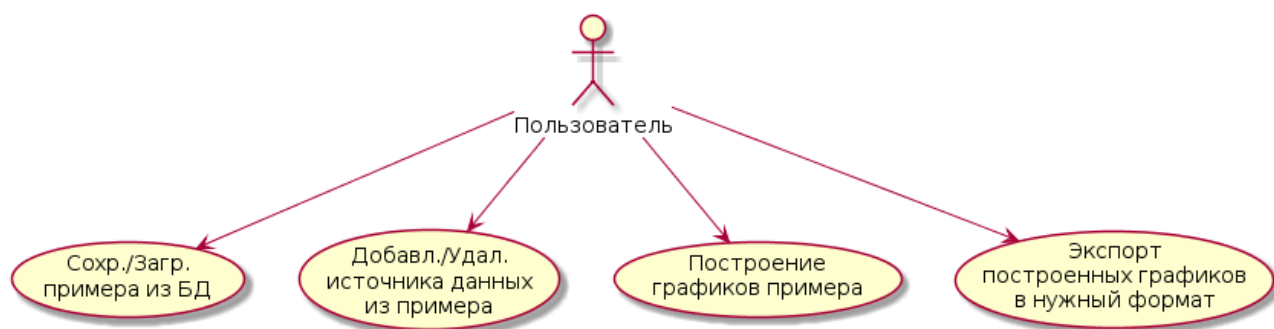


Рисунок 3.1 — Диаграмма прецедентов

В соответствии с рисунком 3.1 представлена базовая функциональность проектируемой программы. Пример или основная сущность программы состоит из списка источников данных. Каждый источник данных – это результат конкретного численного эксперимента. Таким образом достигается цель – работа сразу с несколькими источниками данных.

3.3 Диаграмма классов

Диаграмма классов описывает типы объектов системы и различного рода статические отношения, которые существуют между ними. На диаграммах классов отображаются также свойства классов, операции классов и ограничения, которые накладываются на связи между объектами [9].

Для удобства рассмотрения разобьем диаграмму классов на три рисунка. Каждый из которых соответствует определенной решаемой задаче.

3.3.1 Диаграмма, представляющая способ организации экспериментальных данных

В соответствии с рисунком 3.2 представленная диаграмма классов, иллюстрирует способ организации экспериментальных данных в приложении.

Класс Data – ООП представление данных полученных в результате численного эксперимента в OpenFOAM. Он неразрывно связан с классом-оберткой ExpData – в нем экземпляру присваивается идентификатор и хранится другая служебная информация необходимая программе.

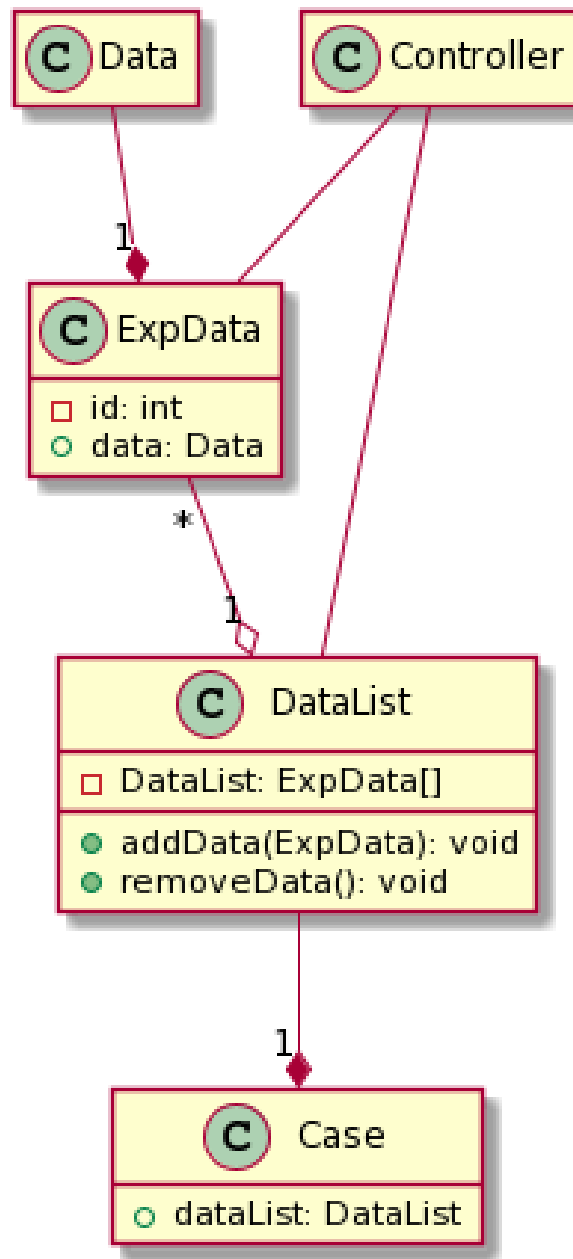


Рисунок 3.2 — Диаграмма классов, представляющая способ организации экспериментальных данных

Так как проектируемое приложение должно работать с некоторым набором экспериментальных данных, то бы введен класс **DataList**, который, как видно в соответствии с рисунком 3.2, содержит список этих данных. Данный класс вступает в качестве поля класса **Case**. Это класс, содержащий основную информацию о текущей сессии приложения, будет рассмотрен чуть более подробно далее.

Класс ExpData и DataList ассоциирован с классом Controller, где происходит их инстанцирование и инициализация. Класс Controller объединяет работу частей модели программы и связывает их с графическим интерфейсом пользователя.

3.3.2 Диаграмма для работы с базой данных и чтения экспериментальных данных

В соответствии с рисунком 3.3 представленная диаграмма классов, иллюстрирует работу приложения с базой данных и чтение экспериментальных данных.

Статический класс ApplicationUtils реализует чтение экспериментальных данных и настроек приложения из файла. Класс DTO (Data Transfer Object) используется в качестве промежуточной сущности для валидации данных.

Класс Database – используется для установления соединения и работы с базой данных. Данный класс будет реализован при помощи паттерна проектирования синглтон (Singleton). Данный паттерн гарантирует, что у класса есть только один экземпляр, и предоставляет к нему глобальную точку доступа [10]. Такой подход дает возможность только один раз при первом обращении произвести ресурсоемкую операцию установления соединения с базой данных, а после каждый раз при последующих обращениях возвращать уже созданный экземпляр класса. Получение различных сущностей из базы данных будет осуществляться при помощи класса DAO (Data Access Object). Этот прием позволит не передавать лишнюю информацию в модель приложения.

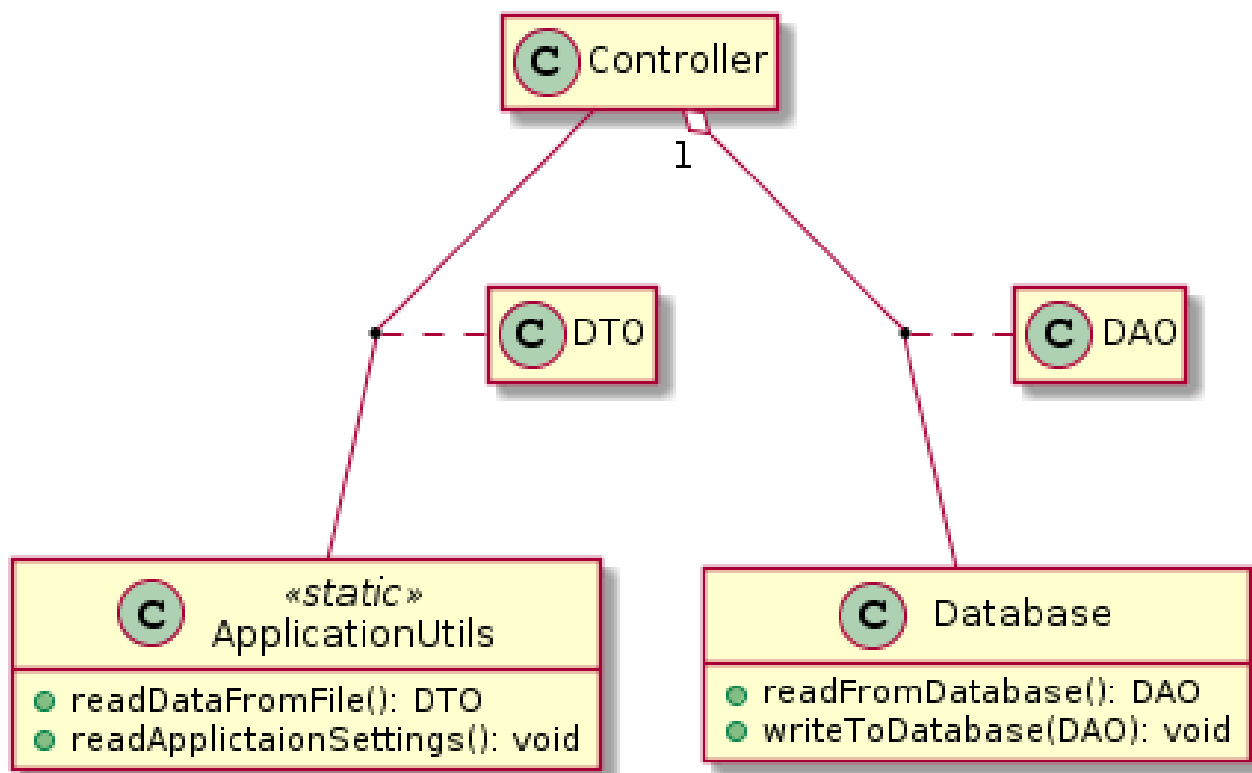


Рисунок 3.3 — Диаграмма классов для работы с базой данных и чтения экспериментальных данных

3.3.3 Диаграмма классов для выполнения операций над графиками

В соответствии с рисунком 3.4 представленная диаграмма классов иллюстрирует организацию классов, реализующих операции над графиками.

Класс *Case* хранит необходимую информацию о текущей сессии, в частности загруженные источники данных (класс *ExpData*) и историю операций – построенные графики. В классе *OperationHistory* непосредственно хранится информация о построенных графиках.

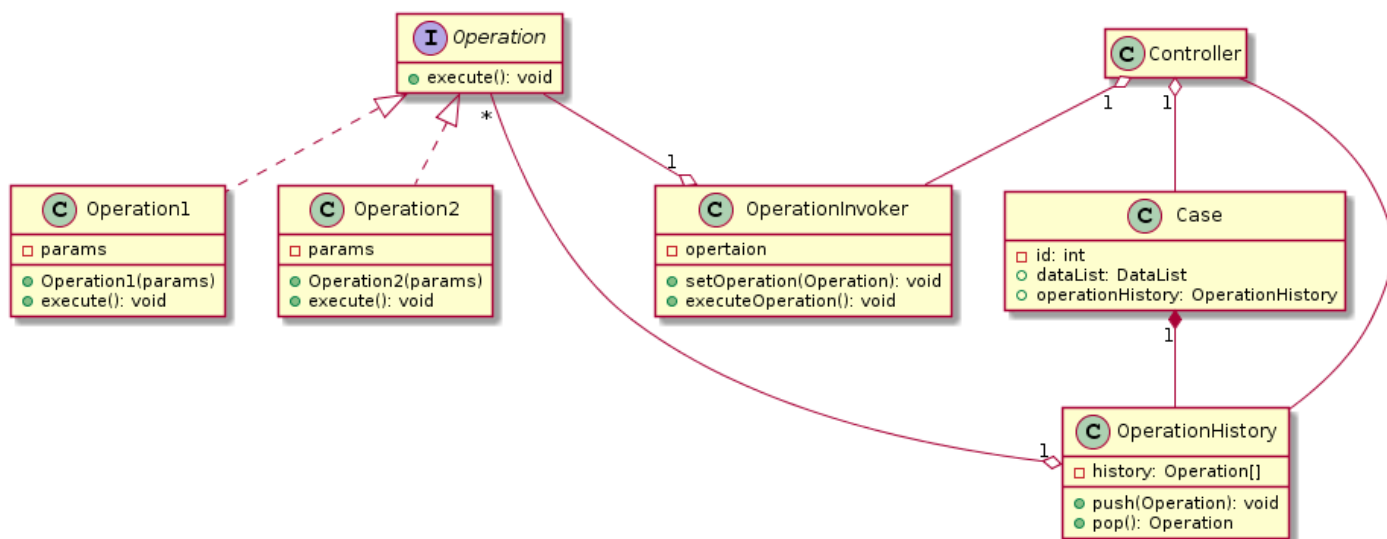


Рисунок 3.4 — Диаграмма классов для выполнения операций над графиками

В целом, существуют разные виды графиков, а пользователем может быть последовательно запланировано несколько различных операций поэтому представляется разумным для организации этой функциональности применить паттерн проектирования команда (command или action).

Команда – это поведенческий паттерн проектирования, который превращает запросы в объекты, позволяя передавать их как аргументы при вызове методов, ставить запросы в очередь, логировать их, а также поддерживать отмену операций [11]. В случае данной информационной системы в роли команды выступает операция построения графика или иная связанная с этим операция.

Структура паттерна следующая:

- Класс `OperationInvoker` хранит ссылку на объект операции и обращается к нему, когда нужно выполнить какое-то действие. `OperationInvoker` работает с операциями только через их общий интерфейс. Он не знает, какую конкретно операцию использует, так как получает готовый объект операции от класса `Controller`.
- `Operation` описывает общий для всех конкретных операций интерфейс. Здесь задан метод `execute` для запуска операции.
- Конкретные операции, это классы `Operation1` и `Operation2`, представленные в соответствии с рисунком 3.4, реализуют различные запросы, следуя общему интерфейсу операций. У них реализованы методы

execute, а конструктор принимает параметры (params), таким образом предоставляя возможность сделать операции неизменяемыми.

- В рамках данного паттерна класс Controller создаёт объекты конкретных операций, передавая в них все необходимые параметры.

3.4 Диаграмма последовательностей

Диаграммы взаимодействия (interaction diagrams) описывают взаимодействие групп объектов в различных условиях их поведения. UML определяет диаграммы взаимодействия нескольких типов, из которых наиболее употребительными являются диаграммы последовательности (sequence diagram) [9].

На диаграмме последовательностей отображаются системные события для одного сценария некоторого прецедента. Поэтому сама диаграмма строится на основе описания прецедента [12].

Если прецедент отвечает на вопрос «Что делает актор?», то последовательность отвечает на вопрос «Как работает система при выполнении данного прецедента?». Каждый прецедент может содержать несколько диаграмм последовательностей, на тот случай, если они описывают несколько альтернативных вариантов развития событий.

Диаграмма последовательностей будет построена только для прецедента «Построение графиков примера», «Сохранение примера в базу данных», «Добавление источника данных».

В соответствии с рисунком 3.5 представлена диаграмма последовательностей для прецедента «Построение графиков примера».

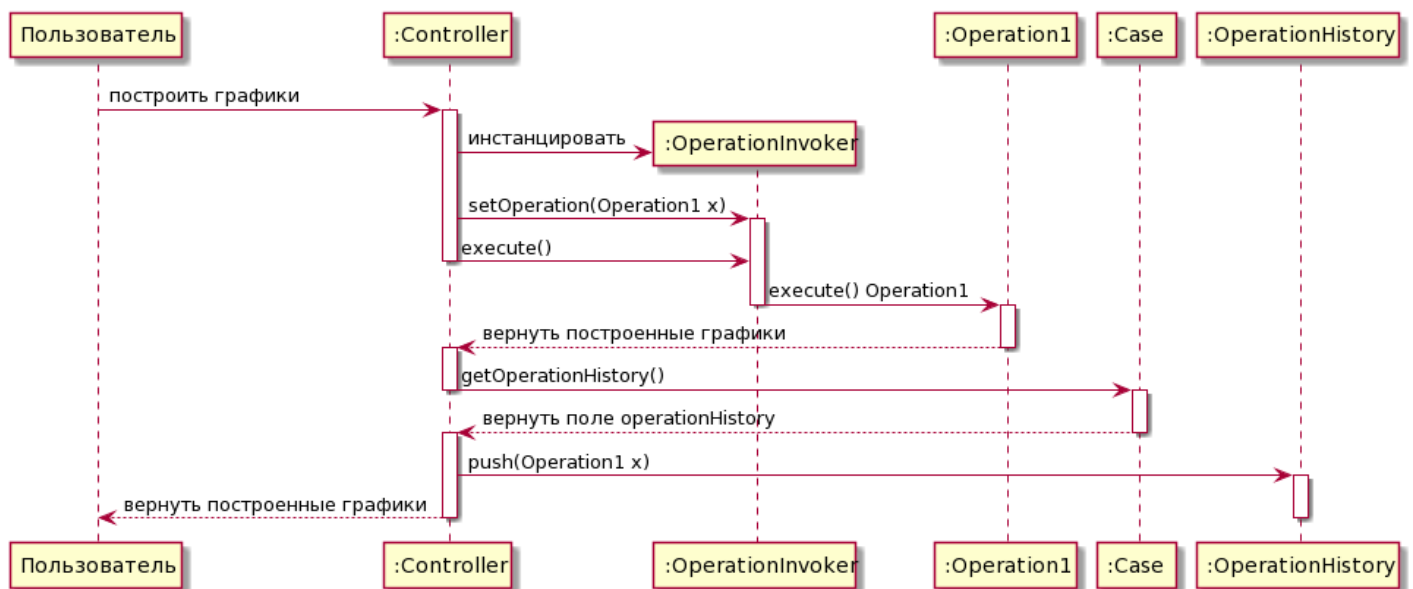


Рисунок 3.5 — Диаграмма последовательностей для прецедента «Построение графиков примера»

В соответствии с рисунком 3.6 представлена диаграмма последовательностей для прецедента «Построение графиков примера».

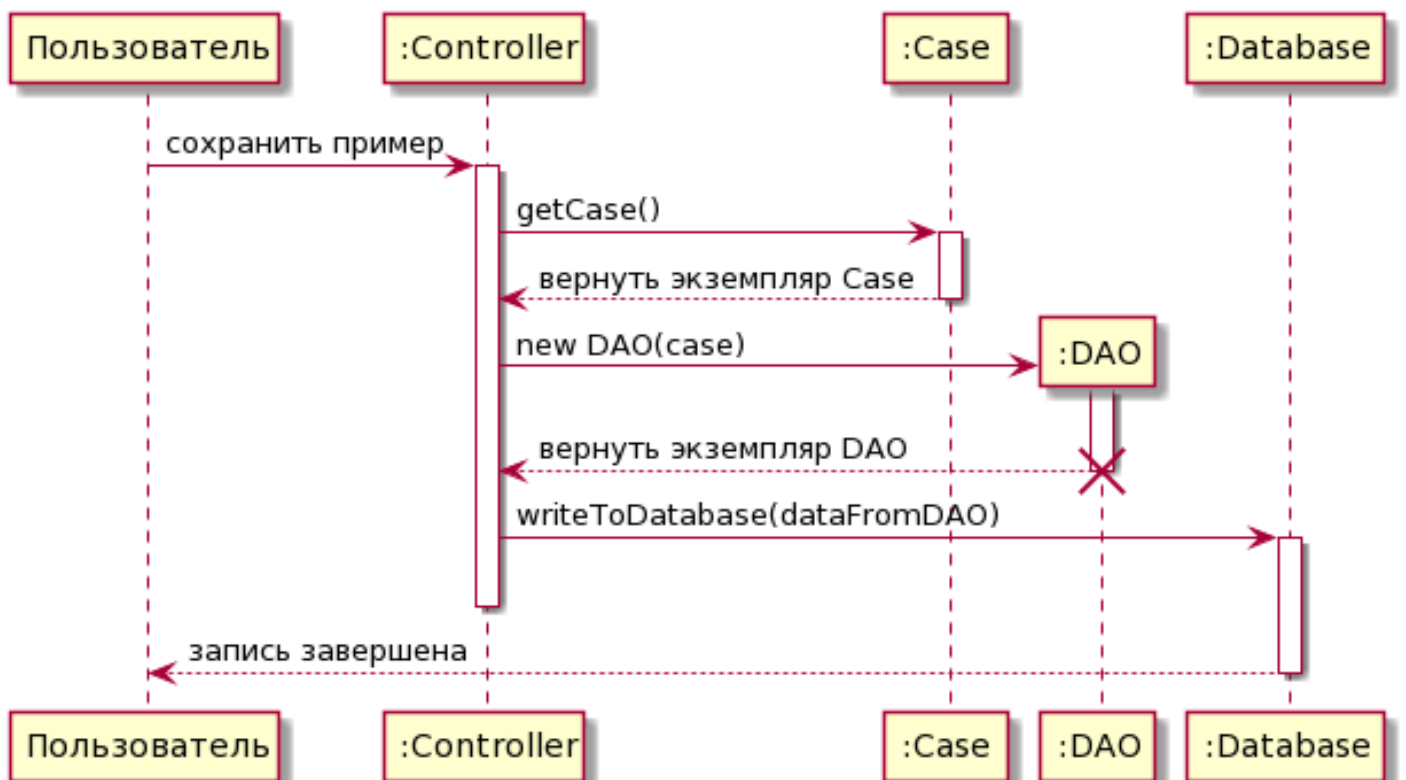


Рисунок 3.6 — Диаграмма последовательностей для прецедента «Сохранение примера в базу данных»

В соответствии с рисунком 3.7 представлена диаграмма последовательностей для прецедента «Добавление источника данных».

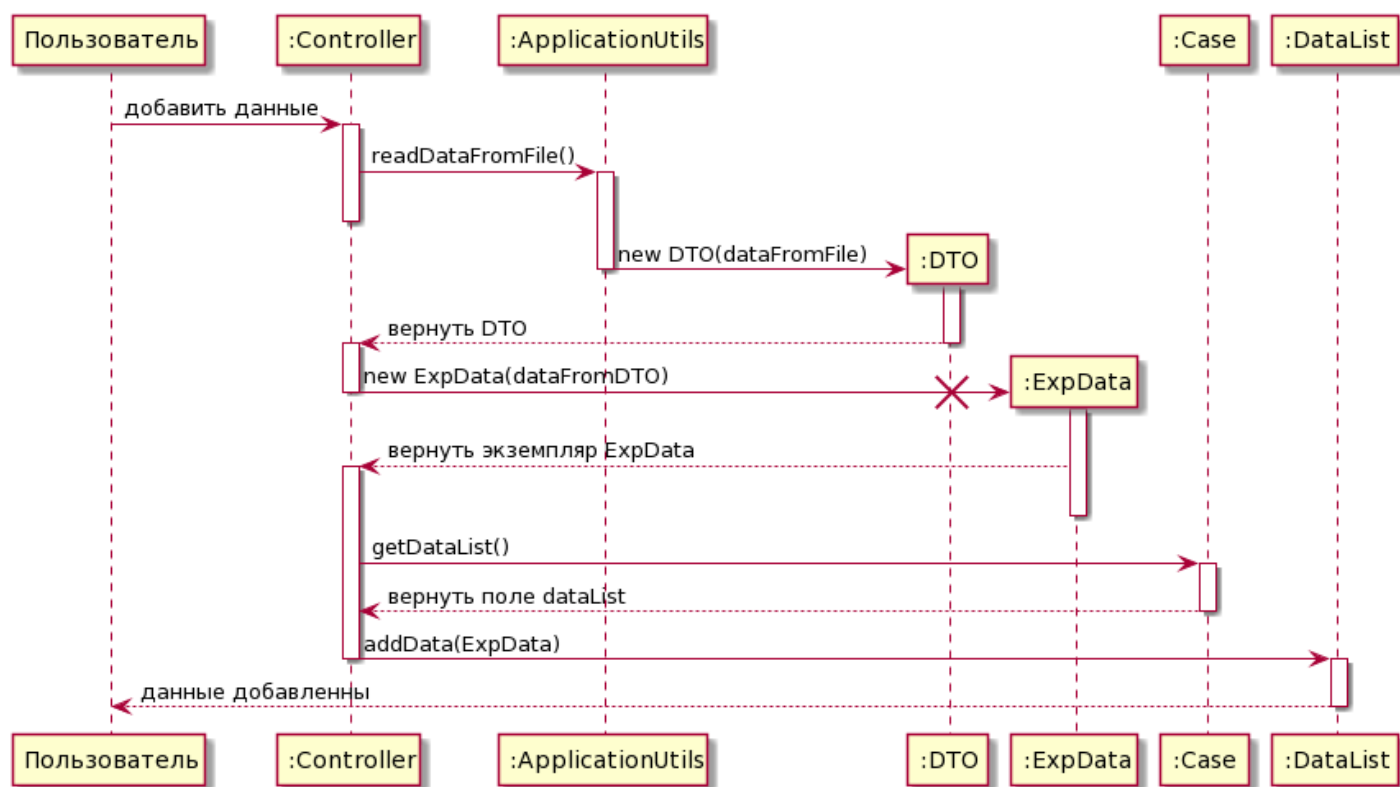


Рисунок 3.7 — Диаграмма последовательностей для прецедента «Добавление источника данных»

4 Выбор средств разработки

Для хранения данных был выбран NoSQL подход. Этот выбор обусловлен теми преимуществами, которые предоставляют СУБД этого вида, а именно:

- Не требуется иметь строгую схему данных
- Линейная масштабируемость

Существует несколько видов NoSQL СУБД:

- База данных «ключ-значение». Представляет собой большую хештаблицу, содержащую пару: ключ и значение. Примеры: Riak, Amazon DynamoDB;
- Документоориентированные СУБД. Хранят документы, состоящие из элементов, каждый из которых отмечен определенным тегом. Примеры: MongoDB, CouchDB;
- Колоночная база данных. В каждом блоке хранятся данные только из одной колонки. Примеры: HBase, Cassandra;
- Графовая база данных. Это сетевая база данных, которая использует узлы и рёбра для отображения и хранения данных. Пример: Neo4J.

В качестве базы данных приложения была выбрана документоориентированная СУБД MongoDB, как одна из наиболее известных СУБД этого вида.

Документоориентированная СУБД работает аналогично базе данных «ключ-значение», однако дополнительно включает в себя кодировку документа. Также она хранит метаданные, связанные с хранимой информацией, что даёт возможность делать запросы на основе этой информации. Учитывая сказанное выше и также то, что такие базы данных работают без схемы данных, делает добавление полей в JSON-документы достаточно простой задачей.

Таким образом, имеется возможность сразу хранить некий аналог объектно-ориентированной модели, что упрощает написание соответствующих классов в программе.

В качестве языка программирования для написания работы был выбран Python, так как он достаточно гибкий и удобный, а также имеет интеграцию с ParaView через библиотеку ParaView Python API и с MongoDB через библиотеку PyMongo.

ЗАКЛЮЧЕНИЕ

В данной работе кратко рассмотрен пакет программ для численного моделирования OpenFOAM, приложения для пост-процессинга ParaView. Была поставлена задача проектирования информационной системы. Выполнена первая итерация проектирования, необходимая для начала написания приложения, в частности построены UML-диаграммы прецедентов, классов, последовательностей. Также были рассмотрены существующие на данном момент решение по пост-обработки экспериментальных данных и выбраны средства разработки.

Таким образом, задачи данной работы были выполнены и, следовательно, поставленная цель была достигнута.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Статья OpenFOAM [Электронный ресурс] : (на 08 апреля 2020 года) // Официальный веб-сайт Wikipedia [Электронный ресурс] : [сайт]. - URL: <https://ru.wikipedia.org/wiki/OpenFOAM> (дата обращения 08.04.2020). - Загл с экрана - Яз. рус.
2. Страница User guide [Электронный ресурс] : (на 08 апреля 2020 года) // Официальный веб-сайт OpenFOAM [Электронный ресурс] : [сайт]. - URL: <https://cfd.direct/openfoam/user-guide> (дата обращения 08.04.2020). - Загл с экрана - Яз. англ.
3. Страница About [Электронный ресурс] : (на 08 апреля 2020 года) // Официальный веб-сайт ParaView [Электронный ресурс] : [сайт]. - URL: <https://www.paraview.org/overview> (дата обращения 08.04.2020). - Загл с экрана - Яз. англ.
4. Статья ParaView [Электронный ресурс] : (на 09 апреля 2020 года) // Официальный веб-сайт Wikipedia [Электронный ресурс] : [сайт]. - URL: <https://ru.wikipedia.org/wiki/ParaView> (дата обращения 09.04.2020). - Загл с экрана - Яз. рус.
5. Страница ParaView and Python [Электронный ресурс] : (на 10 апреля 2020 года) // Официальный веб-сайт ParaView [Электронный ресурс] : [сайт]. - URL: https://www.paraview.org/Wiki/ParaView_and_Python (дата обращения 10.04.2020). - Загл с экрана - Яз. англ.
6. Продукт Helyx OS [Электронный ресурс] : (на 29 марта 2020 года) // Официальный веб-сайт Helyx [Электронный ресурс] : [сайт]. - URL: <https://engys.com/products/helyx-os> (дата обращения 29.03.2020). - Загл с экрана - Яз. англ.
7. Продукт ANSA [Электронный ресурс] : (на 11 апреля 2020 года) // Официальный веб-сайт компании Beta [Электронный ресурс] : [сайт]. - URL: <http://www.beta-cae.gr/ansa.html> (дата обращения 11.04.2020). - Загл с экрана - Яз. англ.

8. Продукт CastNet [Электронный ресурс] : (на 11 апреля 2020 года) // Официальный веб-сайт компании DHCAE Tools [Электронный ресурс] : [сайт]. - URL: <http://www.dhcae-tools.com/CastNet.html> (дата обращения 11.04.2020). - Загл с экрана - Яз. англ.
9. Фаулер, М. UML. Основы, 3-е издание/ М. Фаулер – Спб: Символ-Плюс, 2004. - 192с.
10. Гамма, Э. Приемы объектно-ориентированного проектирования/ Э. Гамма, Р. Хелм, Р. Джонсон, Д. Влиссидес – Спб: Питер, 2019. - 368 с.
11. Паттерны проектирования. Команда [Электронный ресурс] : (на 19 апреля 2020 года) // Веб-сайт refactoring guru [Электронный ресурс] : [сайт]. - URL: <https://refactoring.guru/ru/design-patterns/command> (дата обращения 19.04.2020). - Загл с экрана - Яз. рус.
12. Ларман, К. Применение UML 2.0 и шаблонов проектирования. Практическое руководство. 3-е издание/ К. Ларман – М: И.Д. Вильямс, 2013. - 736 с.