ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук

Департамент программной инженерии

|  |  |
| --- | --- |
| СОГЛАСОВАНО  Научный руководитель  старший преподаватель  департамента больших данных и информационного поиска  факультета компьютерных наук  С.Ю. Папулин  «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2018 г. | УТВЕРЖДЕНО  Академический руководитель  образовательной программы  «Программная инженерия»  профессор департамента программной инженерии, канд. техн. наук  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ В.В. Шилов  «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2018 г. |

|  |  |
| --- | --- |
| Подп. и дата |  |
| Инв. № дубл. |  |
| Взам. инв. № |  |
| Подп. и дата |  |
| Инв. № подл |  |

**БИБЛИОТЕКА ГИСТОГРАММНОЙ МОДЕЛИ ПОД APACHE SPARK**

**Пояснительная записка**

**ЛИСТ УТВЕРЖДЕНИЯ**

**RU.17701729.503100-01 81 01-1-ЛУ**

|  |  |
| --- | --- |
|  | Исполнитель  студент группы БПИ152  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ /П.И. Данилин/  «\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2018 г. |

УТВЕРЖДЕН

RU.17701729.503100-01 81 01-1-ЛУ

**Библиотека гистограммной модели под Apache Spark**

|  |  |
| --- | --- |
| Подп. и дата |  |
| Инв. № дубл. |  |
| Взам. инв. № |  |
| Подп. и дата |  |
| Инв. № подл |  |

**Пояснительная записка**

**RU.17701729.503100-01 81 01-1**

**Листов 42**

**АННОТАЦИЯ**

В данном программном документе приведена пояснительная записка для «Библиотеки гистограммной модели под Apache Spark». Данный программный продукт предназначен для организации вычислений, основанных на гистограммной модели. Он даст возможность пользователю провести анализ массива данных с точки зрения упомянутой модели на кластере компьютеров, совместимых с Apache Spark.

Оформление программного документа «Пояснительная записка» произведено по требованиям ГОСТ 19.404-79 «Пояснительная записка. Требования к содержанию и оформлению»[7].

**Содержание**

[**1.** **ВВЕДЕНИЕ** 6](#_Toc511347080)

[**1.1.** **Наименование разработки** 6](#_Toc511347081)

[**1.2.** **Основания для разработки** 6](#_Toc511347082)

[**2.** **Назначение и область применения** 7](#_Toc511347083)

[**2.1. Назначение программы** 7](#_Toc511347084)

[**2.2. Область применения программы** 7](#_Toc511347085)

[**3.** **Технические характеристики** 8](#_Toc511347086)

[**3.1.** **Постановка задачи для разработки программы** 8](#_Toc511347087)

[**3.2.** **Описание алгоритмов программы** 8](#_Toc511347088)

[**3.2.1.** **Описание алгоритма создания гистограммы** 8](#_Toc511347089)

[**3.2.2.** **Описание алгоритма вычисления операции И** 9](#_Toc511347090)

[**3.2.3.** **Описание алгоритма вычисления операции КРОМЕ** 10](#_Toc511347091)

[**3.2.4.** **Описание алгоритма вычисления операции Вычитание** 11](#_Toc511347092)

[**3.2.5.** **Описание алгоритма вычисления операции Пересечение** 11](#_Toc511347093)

[**3.2.6.** **Описание алгоритма вычисления операции НЕ** 12](#_Toc511347094)

[**3.2.7.** **Описание алгоритма вычисления операции Объединение** 13](#_Toc511347095)

[**3.2.8.** **Описание алгоритма вычисления операции ИЛИ** 14](#_Toc511347096)

[**3.2.9.** **Описание алгоритма вычисления операции Схожесть** 14](#_Toc511347097)

[**3.2.10.** **Описание алгоритма вычисления операции ИСКЛЮЧАЮЩЕЕ КРОМЕ** 15](#_Toc511347098)

[**3.2.11.** **Описание алгоритма вычисления операции ИСКЛЮЧАЮЩЕЕ ИЛИ** 16](#_Toc511347099)

[**3.3.** **Описание и обоснование выбора метода организации входных и выходных данных** 17](#_Toc511347100)

[**3.3.1.** **Описание метода организации входных и выходных данных** 17](#_Toc511347101)

[**3.3.2. Обоснования выбора метода организации входных и выходных данных** 17](#_Toc511347102)

[**3.4.** **Описание и обоснование выбора состава технических и программных средств** 18](#_Toc511347103)

[**3.4.1.** **Состав технических и программных средств** 18](#_Toc511347104)

[**4.** **Ожидаемые технико-экономические показатели** 19](#_Toc511347105)

[**4.1** **Предполагаемая потребность** 19](#_Toc511347106)

[**4.2** **Ориентировочная экономическая эффективность** 19](#_Toc511347107)

[**5.** **Список использованной литературы** 20](#_Toc511347108)

[**ПРИЛОЖЕНИЕ 1** 21](#_Toc511347109)

[**ПРИЛОЖЕНИЕ 2** 22](#_Toc511347110)

[**ПРИЛОЖЕНИЕ 3** 23](#_Toc511347111)

[**ПРИЛОЖЕНИЕ 4** 25](#_Toc511347112)

[**ПРИЛОЖЕНИЕ 5** 26](#_Toc511347113)

# **ВВЕДЕНИЕ**

## **Наименование разработки**

Наименование программного продукта – «Библиотека гистограммной модели под Apache Spark».

Наименование программного продукта на английском языке – «Spark Library for Logic-Set Histogram».

## **Основания для разработки**

Разработка программы ведется на основании приказа Национального исследовательского университета «Высшая Школа Экономики» от 12.12.2017 г. № 2.3-02/1212-01.

# **Назначение и область применения**

## **2.1. Назначение программы**

Программный продукт предназначен для организации распределенных вычислений на основе гистограммной модели с использованием Apache Spark.

## **2.2. Область применения программы**

Данная разработка позволит анализировать большие массивы данных на основе гистограммной модели на компьютерных кластерах, совместимых с Apache Spark, например, под управлением менеджера кластера Hadoop Yarn. Она даст аналитикам и исследователям еще один инструмент обработки коллекций данных с эксплуатацией мощностей кластера.

# **Технические характеристики**

## **Постановка задачи для разработки программы**

Задача для разработки программы состоит в создании программной реализации гистограммной модели в виде библиотеки, которую можно было бы использовать при работе с Apache Spark. Библиотека должна предоставлять следующие функции:

1. Инициализацию элементного состава гистограммной модели;
2. Преобразование определённых типов данных (например, изображений) в гистограммный вид;
3. Формирование элементного высказывания из элементного запроса пользователя;
4. Формирование гистограммного высказывания из элементного с учетом гистограммы данных;
5. Вычисление значения гистограммного высказывания с применением операций[14]: объединения, пересечения, вычитания, И, ИЛИ, исключающего ИЛИ, КРОМЕ, исключающего КРОМЕ, НЕ;
6. Вычисление показателя присутствия элементов из полученного значения гистограммного высказывания;
7. Вычисление показателя схожести между данными по их гистограммам;
8. Применение вычислений для коллекции данных;
9. Сохранение гистограмм данных после преобразования;
10. Сохранение результата вычисления показателя присутствия и схожести.

## **Описание алгоритмов программы**

## **Описание алгоритма создания гистограммы**

Алгоритм создания гистограммы представлен далее (рис. 1).

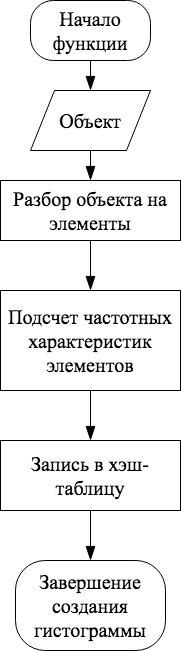


Рисунок 1. Алгоритм создания гистограммы

Сложность: О(количество элементов в объекте)

### **Описание алгоритма вычисления операции И**

Алгоритм вычисления операции И представлен далее (рис. 2).

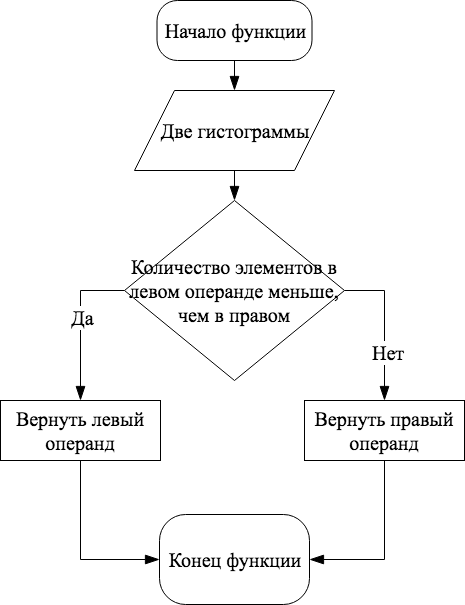


Рисунок 2. Алгоритм вычисления операции И

Входные данные: две гистограммы.

Сложность: О(Количество различных элементов в гистограммах)

### **Описание алгоритма вычисления операции КРОМЕ**

Алгоритм вычисления операции КРОМЕ представлен далее (рис. 3):



Рисунок 3. Алгоритм вычисления операции КРОМЕ

Входные данные: две гистограммы.

Сложность: О(Количество различных элементов в гистограммах)

### **Описание алгоритма вычисления операции Вычитание**

Аналогично операции КРОМЕ (см. п. 3.2.3).

### **Описание алгоритма вычисления операции Пересечение**

Алгоритм вычисления операции Пересечение представлен далее (рис. 4).



Рисунок 4. Алгоритм вычисления операции Пересечение

Входные данные: две гистограммы.

Сложность: О(Количество различных элементов в гистограммах)

### **Описание алгоритма вычисления операции НЕ**

Алгоритм вычисления операции НЕ представлен далее (рис. 5).

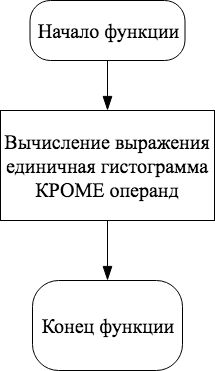


Рисунок 5. Алгоритм вычисления операции НЕ

Входные данные: одна гистограмма.

Сложность: О(операции КРОМЕ)

### **Описание алгоритма вычисления операции Объединение**

Алгоритм вычисления операции Объединение представлен далее (рис. 6).



Рисунок 6. Алгоритм вычисления операции Объединение

Входные данные: две гистограммы.

Сложность: О(Количество различных элементов в гистограммах)

### **Описание алгоритма вычисления операции ИЛИ**

Аналогично операции Объединение (см. п. 3.2.7)

### **Описание алгоритма вычисления операции Схожесть**

Алгоритм вычисления операции Схожесть представлен далее (рис. 7).

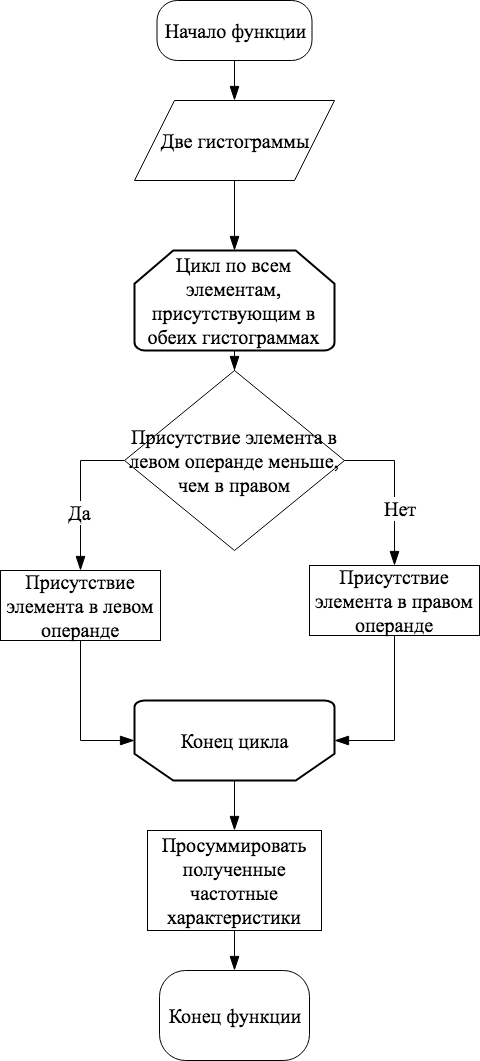


Рисунок 7. Алгоритм вычисления операции Схожесть

Входные данные: две гистограммы.

Сложность: О(Количество различных элементов в гистограммах)

### **Описание алгоритма вычисления операции ИСКЛЮЧАЮЩЕЕ КРОМЕ**

Алгоритм вычисления операции ИСКЛЮЧАЮЩЕЕ КРОМЕ представлен далее (рис. 8).

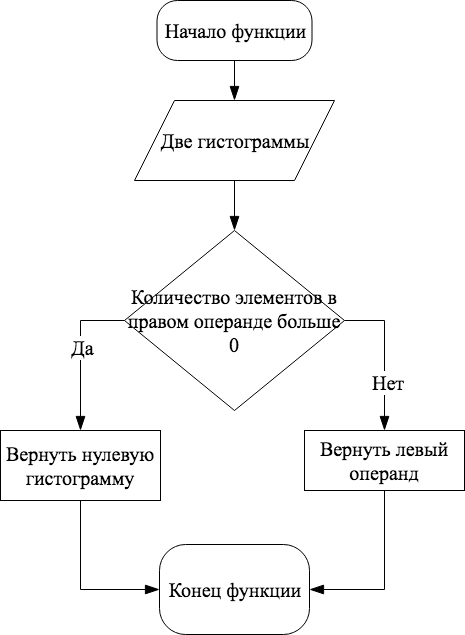


Рисунок 8. Алгоритм вычисления операции ИСКЛЮЧАЮЩЕЕ КРОМЕ

Входные данные: две гистограммы.

Сложность: О(Размер правой гистограммы)

### **Описание алгоритма вычисления операции ИСКЛЮЧАЮЩЕЕ ИЛИ**

Алгоритм вычисления операции ИСКЛЮЧАЮЩЕЕ ИЛИ представлен далее (рис. 9).

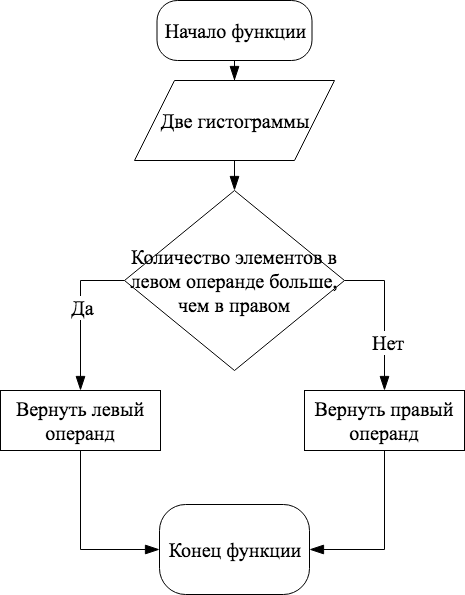


Рисунок 9. Алгоритм вычисления операции ИСКЛЮЧАЮЩЕЕ ИЛИ

Входные данные: две гистограммы.

Сложность: О(Размер правой гистограммы)

## **Описание и обоснование выбора метода организации входных и выходных данных**

### **Описание метода организации входных и выходных данных**

Входные данные должны представлять собой любые объекты, для которых был реализован интерфейс Atomizer для разбора объекта на элементы.

Выходные данные программы представляют собой объекты, реализующие интерфейс Histogram. Также возможна JSON сериализация полученных гистограмм.

### **3.3.2. Обоснования выбора метода организации входных и выходных данных**

Входные данные представляют собой реализацию минимально необходимого для создания гистограммы объектов.

Выходные данные представляют собой реализацию поставленной задачи в виде требуемых объектов гистограмм, либо в JSON сериализованном виде.

## **Описание и обоснование выбора состава технических и программных средств**

### **Состав технических и программных средств**

Для эффективной работы с программой предъявляются следующие требования к составу и параметрам технических средств:

1. 8-ядерный процессор;
2. Оперативная память не менее 8 Гб;
3. 4 жестких диска со свободным объемом не менее 500Гб;
4. Стабильное интернет соединение, желательно 10 Гбит/сек;
5. Монитор;
6. Клавиатура;
7. Мышь.

Для корректной работы программного продукта требуется установленный программный комплекс Apache Spark версии 2.2.0 со всеми зависимостями.

**3.4.2. Обоснование выбора технических и программных средств**

При разработке использовалась IDE Intelij Idea, язык программирования Scala версии 2.11, sbt версии 1.1, библиотеки scalatest для тестирования и json4s для сериализации.

Технические требования связаны с рекомендованными характеристиками для эффективной работы Apache Spark, необходимого по постановке задачи.

# **Ожидаемые технико-экономические показатели**

## **Предполагаемая потребность**

Данная библиотека предоставляет реализацию гистограммной модели, которая послужит в качестве альтернативы существующим методам анализа данных и выявления зависимостей в них, осуществления поиска и обработки. Так как реализаций в данный момент ограниченное количество, данная работа позволит повысить популярность гистограммной модели и даст аналитикам и исследователям новый инструмент.

## **Ориентировочная экономическая эффективность**

Данный инструмент распространяется бесплатно и не требует платы за использование.

# **Список использованной литературы**

1. ГОСТ 19.101-77 Виды программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.102-77 Стадии разработки. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.103-77 Обозначения программ и программных документов. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.104-78 Основные надписи. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.105-78 Общие требования к программным документам. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.201-78 Техническое задание. Требования к содержанию и оформлению. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.603-78 Общие правила внесения изменений. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.604-78 Правила внесения изменений в программные документы, выполненные печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. ГОСТ Р 7.02-2006 Консервация документов на компакт-дисках. Общие требования. – М.: ИПК Издательство стандартов, 2006
11. ГОСТ 18300-87 Спирт этиловый ректификованный технический. Технические условия. – М.:ИПК Издательство стандартов, 1997
12. ГОСТ 9805-84. Спирт изопропиловый. Технические условия. – М.: ИПК Издательство стандартов, 1984.
13. ГОСТ 19.602-78 Правила дублирования, учета и хранения программных документов, выполненных печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001
14. Папулин С.Ю. Анализ коллекции данных посредством логико-множественного гистограммного представления // Программные продукты и системы. 2014. С. 57-60.
15. Apache Spark [Электронный ресурс] // Apache Spark: [сайт]. [2017]. URL: <https:/​/​spark.apache.org>
16. Cистемные требования для Java [Электронный ресурс] // Java: [сайт]. [2017]. URL: <https:/​/​www.java.com/​ru/​download/​help/​sysreq.xml>
17. Jupyter Notebook [Электронный ресурс] // Jupyter: [сайт]. [2018]. URL: https://jupyter.org/
18. Lightweight Scala kernel for Jupyter / IPython 3 [Электронный ресурс] // Github: [сайт]. [2018]. URL: <https://github.com/jupyter-scala/jupyter-scala>
19. Json4s [Электронный ресурс] // Json4s: [сайт]. [2018]. URL: http://json4s.org/

# **ПРИЛОЖЕНИЕ 1**

**Терминология**

**Гистограмма элемента -** модель представления данных, элементы которой упорядочены и соответствуют элементам универсального множества, а их значения указывают на количество этих элементов в данных**.**

**Элементарное высказывание –** высказывание, состоящее из элементов универсального множества и операций между ними.

**Гистограммное высказывание -** высказывание, состоящее из гистограммных элементов и операций между ними.

**Данные** – обрабатываемая гистограммной моделью информация.

**Коллекция –** множество однородных данных.

# **ПРИЛОЖЕНИЕ 2**

**Последовательность операций при формировании и исполнении запроса**

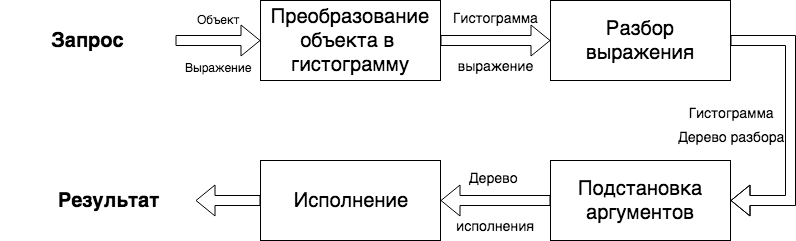
****

Рисунок 1.1. Последовательность операций при формировании и исполнении запроса

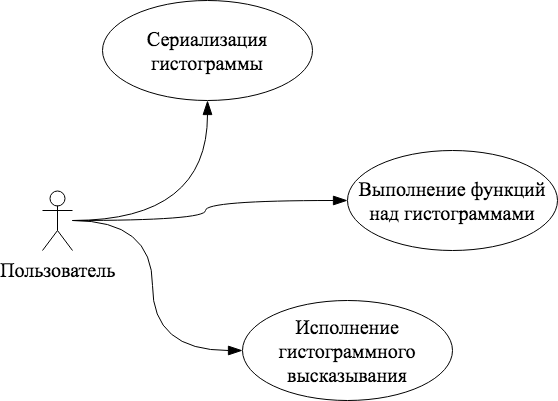


Рисунок 1.2. Диаграмма сценариев использования

# **ПРИЛОЖЕНИЕ 3**

**ДИАГРАММА КЛАССОВ**

Диаграмма классов проекта

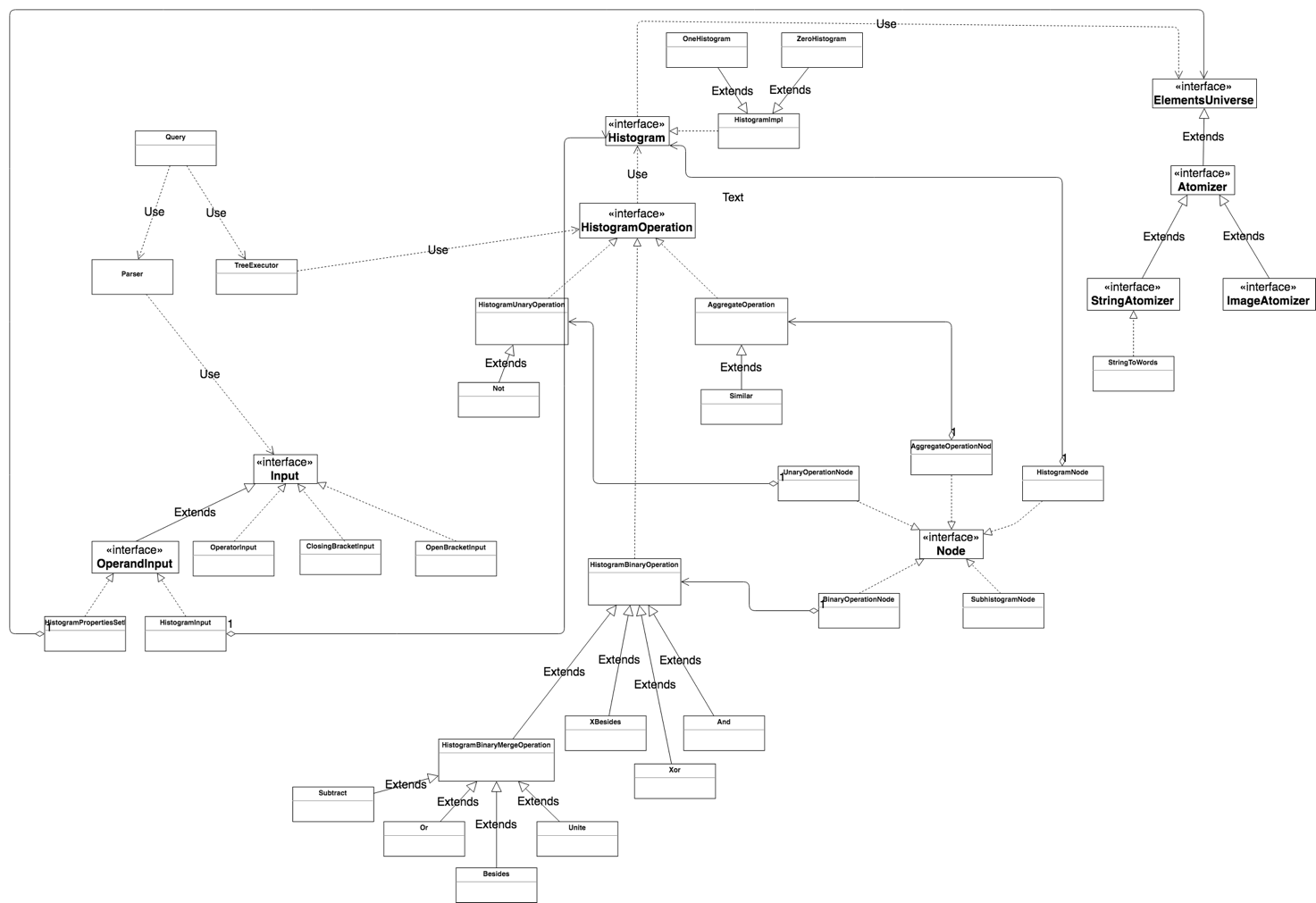
****

Рисунок 1.1. Диаграмма классов проекта

Описание всех сущностей можно найти в приложении 5 данного документа**.**

# **ПРИЛОЖЕНИЕ 4**

**Диаграмма последовательностей**

Диаграмма последовательностей основного юзкейса продукта

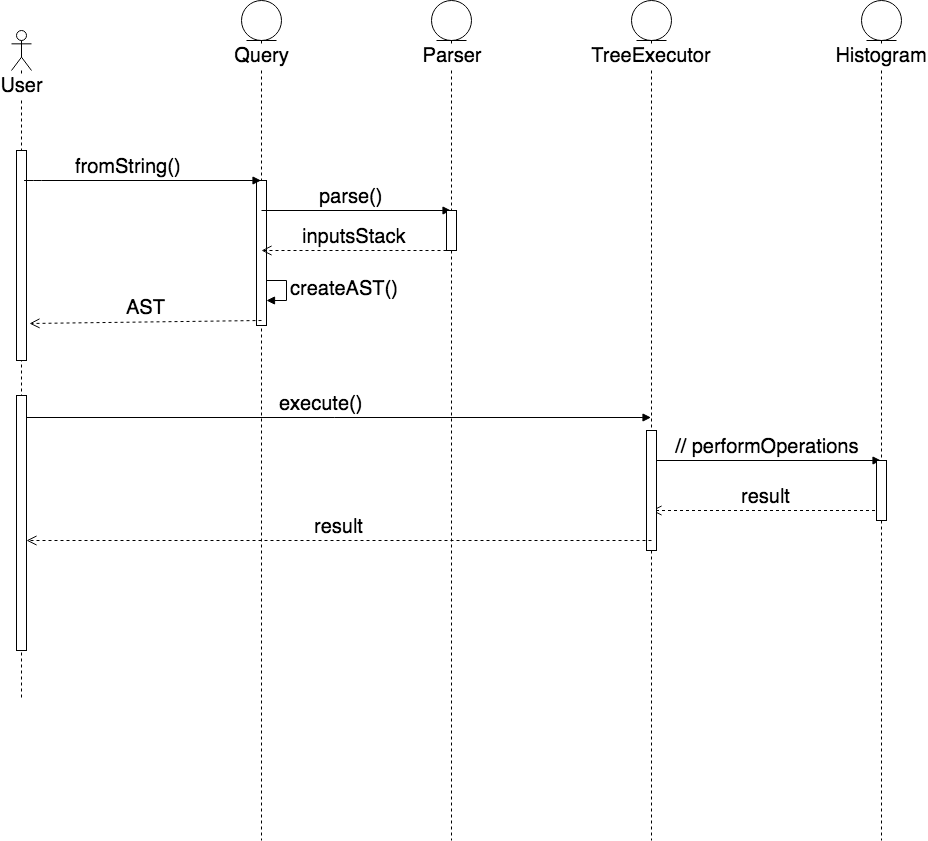
****

Рисунок 1.1. Диаграмма последовательностей основного юзкейса продукта

Описание всех сущностей можно найти в приложении 5 данного документа

# **ПРИЛОЖЕНИЕ 5**

**Описание и функциональное назначение полей и методов**

Таблица 5.1

Описание полей методов и свойств классов

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Класс | | | Atomizer – разбор объектов на элементы | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| atomize | - | Iterable[O] | | source: S | Разбивает входной объект на последовательность элементов универсального множества |
|  |  |  | |  |  |
| Класс | | | ImageToPixels – атомайзер для картинок | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| atomize | - | Iterable[Pixel] | | source: BufferedImage | Разбивает картинку на пиксели: красный, зеленый, синий |
| isElementInUniverse | - | Boolean | | element: Pixel | Есть ли элемент в универсальном множестве |
|  |  |  | |  |  |
| Класс | | | Pixel - цвет | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Pixel | | pixel: Color | Возвращает подходящий тип пикселя к данному |
|  |  |  | |  |  |
| Класс | | | Red – красный цвет | | |
| Класс | | | Green – зеленый цвет | | |
| Класс | | | Blue – синий цвет | | |
| Класс | | | Yellow – желтый цвет | | |
| Класс | | | Magenta – пурпурный цвет | | |
| Класс | | | Cyan – голубой цвет | | |
|  |  |  | |  |  |
| Класс | | | StringAtomizer – атомайзер для строк | | |
|  |  |  | |  |  |
| Класс | | | StringToWords – атомайзер из строк в слова | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| atomize | - | Iterable[Pixel] | | source: String | Разбивает текст по словам |
| isElementInUniverse | - | Boolean | | element: String | Есть ли элемент в универсальном множестве |
|  |  |  | |  |  |
| Класс | | | ElementsUniverse – множество элементов | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| isElementInUniverse | - | Boolean | | element: E | Есть ли элемент в универсальном множестве |
|  |  |  | |  |  |
| Класс | | | Histogram - гистограмма | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Double | | element: E | Частота встречаемости элемента |
| elementsUniverse | - | ElementsUniverse[E] | | - | Универсальное множество гистограммы |
| elementsPresent | - | Set[E] | | - | Элементы, присутствующие в гистограмме |
| subHistogram | - | Histogram[E] | | elements: ElementsUniverse[E] | Гистограмма из поднмножества элементов |
| intersect | - | Histogram[E] | | other: Histogram[E] | Операция пересечения |
| intersect | - | Histogram[E] | | other: ElementsUniverse[E] | Операция пересечения |
| subtract | - | Histogram[E] | | other: Histogram[E] | Операция вычитания |
| subtract | - | Histogram[E] | | other: ElementsUniverse[E] | Операция вычитания |
| - | - | Histogram[E] | | other: Histogram[E] | Операция вычитания |
| - | - | Histogram[E] | | other: ElementsUniverse[E] | Операция вычитания |
| unite | - | Histogram[E] | | other: Histogram[E] | Операция объединения |
| unite | - | Histogram[E] | | other: ElementsUniverse[E] | Операция объединения |
| + | - | Histogram[E] | | other: Histogram[E] | Операция объединения |
| + | - | Histogram[E] | | other: ElementsUniverse[E] | Операция объединения |
| and | - | Histogram[E] | | other: Histogram[E] | Операция И |
| and | - | Histogram[E] | | other: ElementsUniverse[E] | Операция И |
| & | - | Histogram[E] | | other: Histogram[E] | Операция И |
| & | - | Histogram[E] | | other: ElementsUniverse[E] | Операция И |
| besides | - | Histogram[E] | | other: Histogram[E] | Операция кроме |
| besides | - | Histogram[E] | | other: ElementsUniverse[E] | Операция кроме |
| not | - | Histogram[E] | |  | Операция отрицания |
| or | - | Histogram[E] | | other: Histogram[E] | Операция ИЛИ |
| or | - | Histogram[E] | | other: ElementsUniverse[E] | Операция ИЛИ |
| | | - | Histogram[E] | | other: Histogram[E] | Операция ИЛИ |
| | | - | Histogram[E] | | other: ElementsUniverse[E] | Операция ИЛИ |
| xor | - | Histogram[E] | | other: Histogram[E] | Операция ИСКЛЮЧАЮЩЕЕ ИЛИ |
| xor | - | Histogram[E] | | other: ElementsUniverse[E] | Операция ИСКЛЮЧАЮЩЕЕ ИЛИ |
| similar | - | Histogram[E] | | other: Histogram[E] | Операция схожести |
| xbesides | - | Histogram[E] | | other: Histogram[E] | Операция ИСКЛЮЧАЮЩЕЕ КРОМЕ |
| xbesides | - | Histogram[E] | | other: ElementsUniverse[E] | Операция ИСКЛЮЧАЮЩЕЕ КРОМЕ |
| normalize | - | Histogram[E] | |  | Нормализованная гистограмма |
|  |  |  | |  |  |
| Класс | | | HistogramImpl – реализация гистограммы | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| histogram | - | Map[O, Double] | | Словарь входящих элементов в гистограмму в их количество |  |
| universe | - | ElementsUniverse[O] | | Универсальное множество элементов гистограммы |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Double | | element: E | Частота встречаемости элемента |
| elementsUniverse |  | ElementsUniverse[E] | | - | Универсальное множество гистограммы |
| elementsPresent |  | Set[E] | | - | Элементы, присутствующие в гистограмме |
| subHistogram |  | Histogram[E] | | elements: ElementsUniverse[E] | Гистограмма из поднмножества элементов |
| extract | - | Histogram[O] | | source: S)(implicit atomizer: Atomizer[S, O] | Создает гистограмму из объекта |
| normalize | - | Histogram[O] | |  | Нормализованная гистограмма |
|  |  |  | |  |  |
| Класс | | | toHistogramClass – вспомогательный класс для добавления метода toHistogram | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| toHistogram | - | Histogram[O] | | - | Метод для преобразования к гистограмме |
|  |  |  | |  |  |
| Класс | | | SetUniverse – множество элементов и набора элементов | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| unverseSet | - | Set[E] | | Множество, из которого будет строиться универсальное множество |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| isElementInUniverse | - | Boolean | | element: E | Проверяет, есть ли элемент в множестве |
|  |  |  | |  |  |
| Класс | | | OneHistogram – единичная гистограмма | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| universe | - | ElementsUniverse[E] | | Универсальное множество |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Double | | element: E | Частота встречаемости элемента |
| elementsUniverse |  | ElementsUniverse[E] | | - | Универсальное множество гистограммы |
| elementsPresent |  | Set[E] | | - | Элементы, присутствующие в гистограмме |
| subHistogram |  | Histogram[E] | | elements: ElementsUniverse[E] | Гистограмма из поднмножества элементов |
| normalize | - | Histogram[O] | |  | Нормализованная гистограмма |
|  |  |  | |  |  |
| Класс | | | ZeroHistogram – нулевая гистограмма | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| universe | - | ElementsUniverse[E] | | Универсальное множество |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Double | | element: E | Частота встречаемости элемента |
| elementsUniverse |  | ElementsUniverse[E] | | - | Универсальное множество гистограммы |
| elementsPresent |  | Set[E] | | - | Элементы, присутствующие в гистограмме |
| subHistogram |  | Histogram[E] | | elements: ElementsUniverse[E] | Гистограмма из поднмножества элементов |
| normalize | - | Histogram[O] | |  | Нормализованная гистограмма |
|  |  |  | |  |  |
| Класс | | | Operation - операция | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| narity | - | Double | | - | Н-арность операции |
|  |  |  | |  |  |
| Класс | | | HistogramUnaryOperation – унарная операция | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| narity | - | Double | | - | 1 |
| apply | - | Histogram[E] | | Histogram[E] | Применение унарной операции над одним операндом |
|  |  |  | |  |  |
| Класс | | | HistogramBinaryOperation – бинарная операция | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| narity | - | Double | | - | 2 |
| apply | - | Histogram[E] | | first: Histogram[E], second: Histogram[E] | Применение бинарной операции над двумя операндами |
| apply | - | Histogram[E] | | histogram: Histogram[E], properties: ElementsUniverse[E] | Применение бинарной операции над двумя операндами |
|  |  |  | |  |  |
| Класс | | | HistogramBinaryMergeOperation – бинарная операция с поэлементным результатом | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Histogram[E] | | first: Histogram[E], second: Histogram[E] | Применение бинарной операции над двумя операндами |
| merge | - | Double | | leftCount: Double, rightCount: Double | Метод для поэлементного результата операции |
|  |  |  | |  |  |
| Класс | | | And – операция И | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Histogram[E] | | first: Histogram[E], second: Histogram[E] | Применение И над операндами |
| count | - | Double | | histogram: Histogram[E] | Количество элементов в гистограмме |
|  |  |  | |  |  |
| Класс | | | Besides – операция КРОМЕ | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| merge | - | Double | | leftCount: Double, rightCount: Double | Поэлементная логика операции КРОМЕ |
|  |  |  | |  |  |
| Класс | | | Intersect – операция пересечения | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| merge | - | Double | | leftCount: Double, rightCount: Double | Поэлементная логика операции ПЕРЕСЕЧЕНИЕ |
| min | - | Double | | a: Double, b: Double | Минимальное из двух значений |
|  |  |  | |  |  |
| Класс | | | Not – операция НЕ | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Histogram[E] | | histogram: Histogram[E] | Применение операции НЕ |
|  |  |  | |  |  |
| Класс | | | Or – операция ИЛИ | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| merge | - | Double | | leftCount: Double, rightCount: Double | Поэлементная логика операции ИЛИ |
|  |  |  | |  |  |
| Класс | | | Similar – операция схожести | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Histogram[E] | | left: Histogram[E], right: Histogram[E] | Схожесть двух гистограмм |
|  |  |  | |  |  |
| Класс | | | Subtract – операция вычитания | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| merge | - | Double | | leftCount: Double, rightCount: Double | Поэлементная логика операции Вычитание |
|  |  |  | |  |  |
| Класс | | | Unite – операция объединения | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| merge | - | Double | | leftCount: Double, rightCount: Double | Поэлементная логика операции ОБЪЕДИНЕНИЕ |
| max | - | Double | | a: Double, b: Double | Максимальное из двух значений |
|  |  |  | |  |  |
| Класс | | | Xbesides – операция ИСКЛЮЧАЮЩЕЕ КРОМЕ | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Histogram[E] | | left: Histogram[E], right: Histogram[E] | Применение операции ИСКЛЮЧАЮЩЕЕ КРОМЕ |
|  |  |  | |  |  |
| Класс | | | Xor – операция ИСКЛЮЧАЮЩЕЕ ИЛИ | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| apply | - | Histogram[E] | | left: Histogram[E], right: Histogram[E] | Применение операции ИСКЛЮЧАЮЩЕЕ ИЛИ |
|  |  |  | |  |  |
| Класс | | | Node – базовый узел абстрактного синтаксического дерева | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| map | - | Node[E] | | f: Node[E] => Node[E] | Преобразует узел дерева |
|  |  |  | |  |  |
| Класс | | | BinaryOperationNode – узел с бинарной операцией | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| operation | - | HistogramBinaryOperation | | Операция в узле |  |
| left | - | Node[E] | | Левое поддерево |  |
| right | - | Node[E] | | Правое поддерево |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| map | - | Node[E] | | f: Node[E] => Node[E] | Преобразует узел дерева |
|  |  |  | |  |  |
| Класс | | | UnaryOperationNode – узел с унарной операцией | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| operation | - | HistogramUnaryOperation | | Операция в узле |  |
| histogram | - | Node[E] | | Поддерево |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| map | - | Node[E] | | f: Node[E] => Node[E] | Преобразует узел дерева |
|  |  |  | |  |  |
| Класс | | | AggregateOperationNode – узел с операцией, возвращающей Double | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| operation | - | AggregateOperation | | Операция в узле |  |
| left | - | Node[E] | | Левое поддерево |  |
| right | - | Node[E] | | Правое поддерево |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| map | - | Node[E] | | f: Node[E] => Node[E] | Преобразует узел дерева |
|  |  |  | |  |  |
| Класс | | | HistogramNode – узел с гистограммой | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| histogram | - | AggregateOperation | | Гистограмма в узле |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| map | - | Node[E] | | f: Node[E] => Node[E] | Преобразует узел дерева |
|  |  |  | |  |  |
| Класс | | | SubhistogramNode – узел с подгистограммой | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| properties | - | ElementsUniverse[E] | | Подмножество |  |
| originOpt | - | Option[Node[E]] | | Поддерево узла |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| map | - | Node[E] | | f: Node[E] => Node[E] | Преобразует узел дерева |
|  |  |  | |  |  |
| Класс | | | OperationInput – ввод операции | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| operation | - | Operation | | Вводимая операция |  |
|  |  |  | |  |  |
| Класс | | | SubhistogramInput – ввод взятия подгистограммы | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| properties | - | ElementsUniverse[E] | | Подмножество элементов |  |
|  |  |  | |  |  |
| Класс | | | HistogramInput – ввод гистограммы | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| histogram | - | Histogram[E] | | Гистограмма |  |
|  |  |  | |  |  |
| Класс | | | Parser – парсер строки для запроса | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| parse | - | Option[Stack[Input[E]]] | | query: String, implicit aliasToInput: Map[String, Input[E]] | Преобразует строку в последовательность входных лексем в польской нотации |
| getLexems | - | Option[Stack[Input[E]]] | | query: String, acc: Stack[Input[E]], aliasToInput: Map[String, Input[E]] | Преобразует строку в последовательность входных конструкций |
| toPolishNotation | - | Stack[Input[E]] | | query: Stack[Input[E]], resultAcc: Stack[Input[E]], operandsAcc: Stack[Input[E]] | Преобразует последовательность входов в польскую инверсную последовательность |
|  |  |  | |  |  |
| Класс | | | TreeExecutor – выполняет операции над АСД | | |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| execute | - | Either[Histogram[E], Double] | | tree: Node[E] | Исполняет вычисления на АСТ |
|  |  |  | |  |  |
| Класс | | | Query - запрос | | |
| Поля | | | | | |
| Имя | Модификатор доступа | Тип | | Назначение |  |
| root | - | Node[E] | | АСТ |  |
| standardAliases | - | Map[String, Input[E]] | | Стандартные алиасы входных конструкций |  |
| Методы | | | | | |
| Имя | Модификатор доступа | Тип | | Аргументы | Назначение |
| execute | - | Either[Histogram[E], Double] | | histogram: Histogram[E] | Исполняет запрос |
| parseStack | - | Node[E] | | operationsStack: Stack[Input[E]] | Преобразует польскую запись в АСТ входных конструкций |
| apply | - | Query[E] | | Stack[Input[E]] | Формирует запрос по последовательности входных конструкций |
| fromString | - | Query[E] | | query: String, aliasToInput: Map[String, Input[E]] | Формирует запрос по строковому представлению |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ** | | | | | | | | | |
| Изм. | Номера листов (страниц) | | | | Всего  листов  (страниц)  в докум. | № докум. | Входящий № сопроводитель-ного документа и дата | Подпись | Дата |
| изме-ненных | заме-  ненных | новых | анну-  лиро-  ванных |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |