**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** |  |

**РЕАЛИЗАЦИЯ БИБЛИОТЕКИ ГИСТОГРАММНОЙ МОДЕЛИ ПОД APACHE SPARK**

**Программа и методика испытаний**

**ЛИСТ УТВЕРЖДЕНИЯ**

**RU.17701729.503100-01 51 01-1-ЛУ**

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

УТВЕРЖДЕН

RU.17701729.503100-01 51 01-1-ЛУ

**РЕАЛИЗАЦИЯ БИБЛИОТЕКИ ГИСТОГРАММНОЙ МОДЕЛИ ПОД APACHE SPARK**

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

**Программа и методика испытаний**

**RU.17701729.503100-01 51 01-1**

**Листов 14**

# **АННОТАЦИЯ**

В данном программном документе приведена программа и методика испытаний для «Реализация библиотеки гистограммной модели под Apache Spark».

Оформление программного документа «Программа и методика испытаний» произведено по требованиям ГОСТ 19.301-79 «Программа и методика испытаний. Требования к содержанию и оформлению»[7].

**Содержание**

[**АННОТАЦИЯ** 2](#_Toc511235084)

[**1.** **Объект испытаний** 4](#_Toc511235085)

[**2.** **Цель испытаний** 4](#_Toc511235086)

[**3.** **Требования к программе** 5](#_Toc511235087)

[**4.** **Требования к программной документации** 5](#_Toc511235088)

[**4.1. Состав программной документации** 5](#_Toc511235089)

[**5.** **Средства и порядок испытаний** 6](#_Toc511235090)

[**5.1** **Технические средства** 6](#_Toc511235091)

[**5.2** **Программные средства** 6](#_Toc511235092)

[**6.** **Методы испытаний** 6](#_Toc511235093)

[**6.1** **Испытание cоздания объекта гистограммной модели** 6](#_Toc511235094)

[**6.2** **Испытание преобразования картинок и текстов в гистограммный вид** 7](#_Toc511235095)

[**6.3** **Испытание формирования элементного высказывания из элементного запроса пользователя** 7](#_Toc511235096)

[**6.4** **Формирование гистограммного высказывания из элементного с учетом гистограммы данных** 7](#_Toc511235097)

[**6.5** **Вычисление значения гистограммного высказывания с применением операций: объединения, пересечения, вычитания, И, ИЛИ, исключающего ИЛИ, КРОМЕ, исключающего КРОМЕ, НЕ** 8](#_Toc511235098)

[**6.6** **Вычисление показателя присутствия элементов из полученного значения гистограммного высказывания** 9](#_Toc511235099)

[**6.7** **Вычисление показателя схожести между данными по их гистограммам** 9](#_Toc511235100)

[**6.8** **Сохранение гистограмм данных после преобразования** 10](#_Toc511235101)

[**6.9** **Сохранение результата вычисления показателя присутствия и схожести** 10](#_Toc511235102)

[**Список использованной литературы** 11](#_Toc511235103)

[**Приложение 1** 12](#_Toc511235104)

# **Объект испытаний**

Объектом испытаний является программа «Реализация библиотеки гистограммной модели под Apache Spark».

# **Цель испытаний**

Целью проведения испытаний является проверка правильности выполнения программой функций, изложенных в разделе «Требования к программе».

# **Требования к программе**

Программа должна проходить проверку функциональных требований, указанных в документе «Реализация библиотеки гистограммной модели под Apache Spark». Техническое задание.

Программа должна обеспечивать возможность выполнения перечисленных ниже функций:

1. Создание объекта 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. Сохранение гистограмм данных после преобразования
9. Сохранение результата вычисления показателя присутствия и схожести

# **Требования к программной документации**

## **4.1. Состав программной документации**

1. «Реализация библиотеки гистограммной модели под Apache Spark». Техническое задание. ГОСТ 19.71-208
2. «Реализация библиотеки гистограммной модели под Apache Spark» Пояснительная записка. ГОСТ 19.404-79
3. «Реализация библиотеки гистограммной модели под Apache Spark» Программа и методика испытаний. ГОСТ 19.301-79
4. «Реализация библиотеки гистограммной модели под Apache Spark» Текст программы ГОСТ 19.401-78

# **Средства и порядок испытаний**

* 1. **Технические средства**

Компьютер должен обладать следующими характеристиками и периферией:

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
   1. **Программные средства**

Для корректной работы программного продукта требуется установленный программный комплекс Apache Spark версии 2.2.0 со всеми зависимостями, а также установленный программный комплекс jupyter[17] со scala kernel[18].

Помимо данных средств потребуются файлы из папки Demonstration на диске с файлами проекта.

**5.3. Порядок проведения испытаний**

Для установки проекта требуется скопировать содержимое папки Test и запустить файл Demonstration.ipynb с помощью jupyter.

1. **Методы испытаний**

Открываем файл Demonstration.ipynb с помощью jupyter.

## **Испытание cоздания объекта гистограммной модели**

Запустим первые 3 ячейки, убедимся по выводу под 3 ячейкой, что гистограмма создалась правильно.



Рисунок 1. Создание объекта гистограммной модели

## **Испытание** **преобразования картинок и текстов в гистограммный вид**

Создадим гистограмму на основе тестового текста и убедимся, что она правильно сконструировалась:

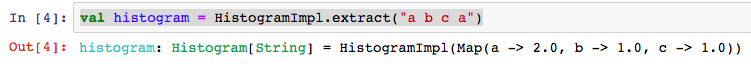


Рисунок 2. Преобразование картинок и текстов в гистограммный вид

## **Испытание** **формирования элементного высказывания из элементного запроса пользователя**

Исполним вспомогательные ячейки до ячейки с кодом:

val query = Query.fromString("red|yellow")(aliases)

val result = query.execute(imagesHists("tiger")).left.get

result.draw

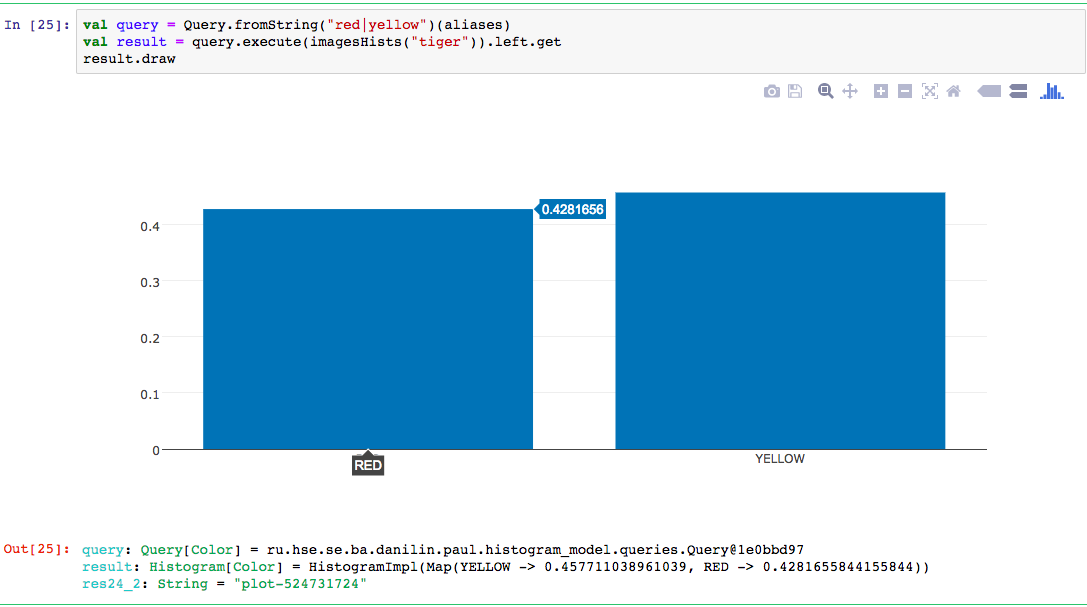


Рисунок 3. Формирование элементного высказывания из элементного запроса пользователя

Убедимся, что была нарисована гистограмма.

## **Формирование гистограммного высказывания из элементного с учетом гистограммы данных**

Исполним ячейку с кодом:

val newQuery = new Query(query.injectHistogram(imagesHists("sea")))

val result = newQuery.execute(imagesHists("tiger")).left.get

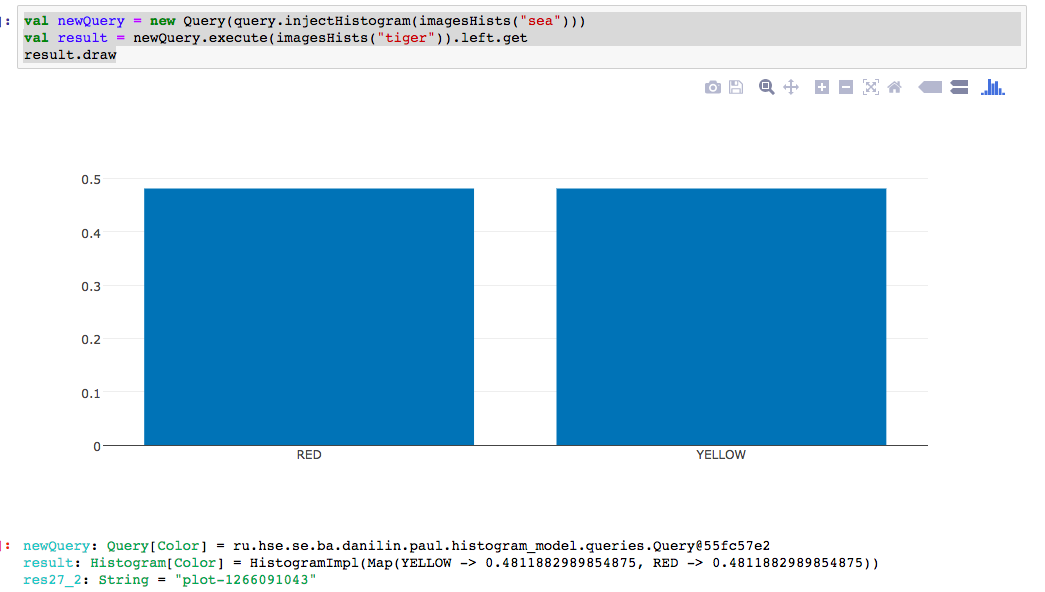
result.draw

Рисунок 4. Формирование гистограммного высказывания из элементного с учетом гистограммы данных

Убедимся, что гистограмма получилась другая относительно предыдущего пункта.

## **Вычисление значения гистограммного высказывания с применением операций: объединения, пересечения, вычитания, И, ИЛИ, исключающего ИЛИ, КРОМЕ, исключающего КРОМЕ, НЕ**

Исполним ячейку со следующим кодом:

val result = imagesHists("sea") + imagesHists("tiger")

result.draw

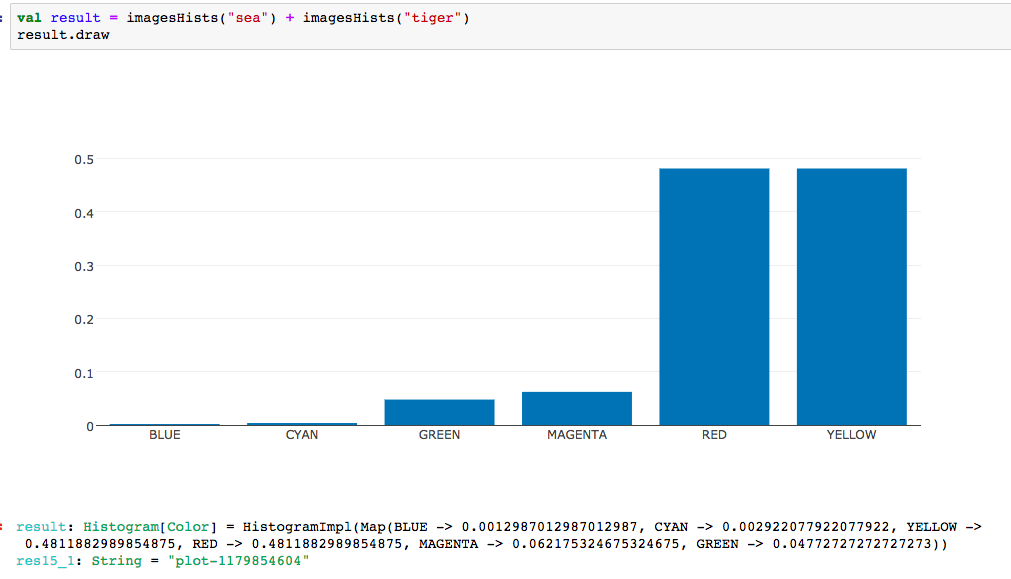


Рисунок 5. Вычисление значения гистограммного высказывания с применением операций

Убедимся, что отрисованная диаграмма соответствует ожиданиям.

## **Вычисление показателя присутствия элементов из полученного значения гистограммного высказывания**

Исполним ячейку со следующим кодом:

result(Blue)

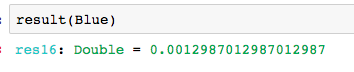


Рисунок 6. Вычисление показателя присутствия элементов из полученного значения гистограммного высказывания

Убедимся, что выведенное число соответствует присутствию синего в гистограмме.

## **Вычисление показателя схожести между данными по их гистограммам**

Исполним ячейку со следующим кодом:

imagesHists("sea") similar imagesHists("tiger")

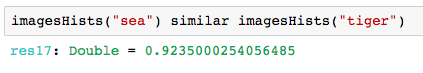


Рисунок 7. Вычисление показателя схожести между данными по их гистограммам

Проверим, что результат показателя схожести соответствует ожиданиям.

## **Сохранение гистограмм данных после преобразования**

Исполним ячейку со следующим кодом:

import ru.hse.se.ba.danilin.paul.histogram\_model.histogram.Implicits.\_

import org.json4s.native.Serialization.{write, read => readJson}

val json = write(result)

val writer = new PrintWriter(new File("test result.json"))

writer.write(json)

writer.close()

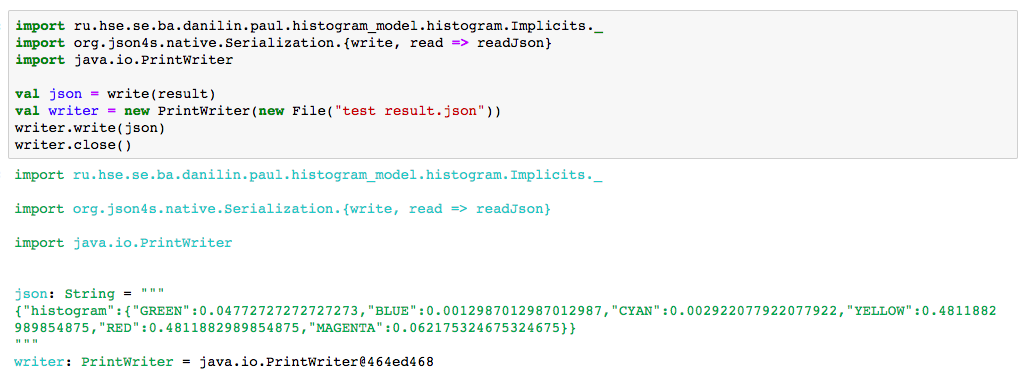


Рисунок 8. Сохранение гистограмм данных после преобразования

Проверим, что файл сохранился.

## **Сохранение результата вычисления показателя присутствия и схожести**

Исполним ячейку со следующим кодом:

val writer = new PrintWriter(new File("test result.json"))

writer.write(imagesHists("sea") similar imagesHists("tiger"))

writer.close()



Рисунок 9. Сохранение результата вычисления показателя присутствия и схожести

Проверим, что файл сохранился.

# **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.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **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 |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |