


ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Департамент программной инженерии

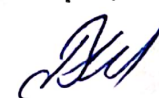
СОГЛАСОВАНО

Научный руководитель,
старший преподаватель департамента
программной инженерии факультета
компьютерных наук


«22» мая 2019 г. А.В. Меликян

УТВЕРЖДАЮ

Академический руководитель
образовательной программы
«Программная инженерия»
профессор департамента программной
инженерии, канд. техн. наук


«22» 05 2019 г. В.В. Шилов

ПРОГРАММА ДЛЯ КЛАСТЕРИЗАЦИИ РОССИЙСКИХ ВУЗОВ ПО
ПОКАЗАТЕЛЯМ ИХ НАУЧНО-ОБРАЗОВАТЕЛЬНОЙ ДЕЯТЕЛЬНОСТИ
НА ОСНОВЕ ИЕРАРХИЧЕСКОГО АГЛОМЕРАТИВНОГО МЕТОДА

Руководство оператора

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729.04.13-01 34 01-1-ЛУ

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

Исполнитель
студент группы БПИ181
/А.А. Матевосян /
«22» май 2019 г.

Москва 2019

УТВЕРЖДЕН
RU.17701729.04.13-01 34 01-1-ЛУ

**ПРОГРАММА ДЛЯ КЛАСТЕРИЗАЦИИ РОССИЙСКИХ ВУЗОВ ПО
ПОКАЗАТЕЛЯМ ИХ НАУЧНО-ОБРАЗОВАТЕЛЬНОЙ ДЕЯТЕЛЬНОСТИ
НА ОСНОВЕ ИЕРАРХИЧЕСКОГО АГЛОМЕРАТИВНОГО МЕТОДА**

Руководство оператора

RU.17701729.04.13-01 34 01-1

Листов 28

<i>Подп. и дата</i>	
<i>Инв. № дубл.</i>	
<i>Взам. инв. №</i>	
<i>Подп. и дата</i>	
<i>Инв. № подл</i>	

Москва 2019

АННОТАЦИЯ

Руководство оператора – это документ, назначение которого — предоставить людям помощь в использовании некоторого программного продукта.

Настоящее Руководство оператора предназначено для правильной организации работы с «Программы для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода».

Руководство оператора для «Программы для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода» содержит следующие разделы: «Назначение программы», «Условия выполнения программы», «Выполнение программы», «Сообщения оператору» и приложения [7].

В разделе «Назначение программы» указаны сведения о назначении программы и информация о функциях и принципе эксплуатации программы.

Раздел «Условия выполнения программы» содержит информацию об условиях, необходимых для выполнения данной программы (минимальный состав аппаратурных и программных средств).

Раздел «Выполнение программы» содержит последовательность действий оператора, обеспечивающих загрузку, запуск, выполнение и завершение программы, описание функций, формата и возможных вариантов команд, с помощью которых оператор осуществляет загрузку и управляет выполнением программы, а также ответы программы на эти команды.

В разделе «Сообщения оператору» указаны тексты сообщений, выдаваемые в ходе выполнения программы, описание их содержания и соответствующие действия оператора.

Настоящий документ разработан в соответствии с требованиями:

- 1) ГОСТ 19.101-77 Виды программ и программных документов [1];
- 2) ГОСТ 19.102-77 Стадии разработки [2];
- 3) ГОСТ 19.103-77 Обозначения программ и программных документов [3];
- 4) ГОСТ 19.104-78 Основные надписи [4];
- 5) ГОСТ 19.105-78 Общие требования к программным документам [5];
- 6) ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом [6];
- 7) ГОСТ 19.505-79 Руководство оператора. Требования к содержанию и оформлению [7].

Изменения к данному Руководству оператора оформляются согласно ГОСТ 19.604-78 [8], ГОСТ 19.603-78 [9].

Перед прочтением данного документа рекомендуется ознакомиться с терминологией, приведенной в Приложении 1 настоящего технического задания.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

СОДЕРЖАНИЕ

1.	НАЗНАЧЕНИЕ РАЗРАБОТКИ	5
1.1.	Функциональное назначение.....	5
1.2.	Эксплуатационное назначение.....	5
1.3.	Состав функций	5
2.	УСЛОВИЯ ВЫПОЛНЕНИЯ ПРОГРАММЫ.....	7
2.1.	Минимальный состав аппаратурных средств	7
2.2.	Минимальный состав программных средств.....	7
2.3.	Требования к персоналу (пользователю)	7
3.	ВЫПОЛНЕНИЕ ПРОГРАММЫ	8
3.1.	Установка и запуск программы.....	8
3.2.	Стартовая страница	10
3.3.	Подготовка программы к работе.....	11
3.4.	Работа с входными данными	11
3.1.	Открытие файла	11
3.2.	Закрытие файла	12
3.3.	Сохранение файла.....	12
3.5.	Работа с таблицей данных	13
3.1.	Добавление новой строки	13
3.2.	Изменение данных в таблице	13
3.3.	Удаление строки.....	13
3.4.	Фильтрация данных	13
3.5.	Сортировка данных.....	13
3.6.	Кластеризация данных	13
3.1.	Выполнение кластеризации данных.....	13
3.2.	Построить дендограмму	14
3.3.	Показать описательную статистику кластеров	15
3.7.	Экспорт.....	15
3.8.	О программе.....	16
3.9.	Завершение работы с программой	16
4.	СООБЩЕНИЯ ОПЕРАТОРУ	17
5.	СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	19
	ПРИЛОЖЕНИЕ 1 - ТЕРМИНОЛОГИЯ	20
	ПРИЛОЖЕНИЕ 2 - ДИАГРАММА ПРЕЦЕДЕНТОВ	21
	ПРИЛОЖЕНИЕ 3 - ПРИМЕР КОНФИГУРАЦИОННОГО ФАЙЛА DATACONFIG.XML	22
	ПРИЛОЖЕНИЕ 4 - ПРИМЕР ВХОДНОГО ФАЙЛА	23
	ПРИЛОЖЕНИЕ 5 - ПРИМЕР ВЫХОДНОГО ФАЙЛА ТАБЛИЦЫ КЛАСТЕРОВ	24
	ПРИЛОЖЕНИЕ 6 - ПРИМЕР ВЫХОДНОГО ФАЙЛА ОПИСАТЕЛЬНОЙ СТАТИСТИКИ КЛАСТЕРОВ.....	25
	ПРИЛОЖЕНИЕ 7 - ПРИМЕР ВЫХОДНОГО ФАЙЛА КАРТИНКИ ДЕНДОГРАММЫ ...	26

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ.....27

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

1. НАЗНАЧЕНИЕ РАЗРАБОТКИ

1.1. Функциональное назначение

Данная программа является инструментом для кластерного анализа данных используя агломеративный метод иерархической кластеризации:

- Кластеризация данных по выбранным показателям
- Построения дерева кластеров из исходных данных
- Построение таблицы объектов с описанием кластеров, которым данные объекты принадлежат
- Нормализацию данных
- Показать статистическую таблицу данных из полученных кластеров

1.2. Эксплуатационное назначение

Кластерный анализ является одним из ведущих направлений в сфере описательной статистики и машинного обучения. И реализации решении задач кластеризации часто используются в быту.

Например, кластерный анализ может использоваться в следующих сферах:

- В качестве метода для классификации различных объектов
- В качестве метода для группирования некоторых объёмных запросов в интернете в более компактные кластеры для дальнейшей обработки
- В качестве метода для нахождения закономерностей в данных
- В качестве метода для генерации подобных данных
- В качестве метода для обработки больших данных и разделение их на категории для дальнейшей обработки

1.3. Состав функций

Программа должна обеспечивать возможность выполнения перечисленных ниже функций:

- Уметь открывать входные данные и сохранять их
- Предоставить данные в виде таблицы
- Иметь возможность добавления новой строки и возможность изменять и удалять строки из таблицы
- Фильтровать строки в таблице по заданному значения фильтрации используя выбранный пользователем столбец и условие фильтрации
- Сортировать выбранный пользователем столбец в алфавитном порядке по возрастанию или по убыванию
- Иметь возможность отмены последнего действия и повтора последнего действия (в качестве действия подразумевается выполнение фильтрации и сортировки)
- Кластеризовать данные по выбранному числу кластеров, меры расстояния, стратегии объединения и метода нормализации
- Рассчитать число рекомендуемых кластеров по индексу Цалиньски-Харабаша
- Построить дендограмму из полученного дерева кластеров

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

- Показать описательную статистику по среднему значению данных в кластере
- Показать таблицу с значением искомого кластера для объекта
- Сохранять дендограмму, таблицу кластеров и таблицу описательной статистики кластеров

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

2. УСЛОВИЯ ВЫПОЛНЕНИЯ ПРОГРАММЫ

2.1. Минимальный состав аппаратных средств

Для надёжной и бесперебойной работы программы требуется следующий состав технических средств [14]:

- 1) персональный компьютер, оснащенный 32-разрядным (x86) или 64-разрядным (x64) процессором с тактовой частотой 1 ГГц и выше
- 2) 1 ГБ для x86 и 2 ГБ для x64 оперативной памяти или больше
- 3) не менее 16 ГБ для x86 и 20 ГБ для x64 свободного места на жестком диске
- 4) видеокарта и монитор с разрешением не менее чем 1366x768 точек
- 5) мышь или совместимое указывающее устройство
- 6) клавиатура

2.2. Минимальный состав программных средств

Для работы программы необходим следующий состав программных средств:

- 1) операционная система Microsoft Windows 7 SP1 или более поздняя версия;
- 2) установленный Microsoft .NET Framework 4.7.1, требующий Windows Installer 5.0 или более поздняя версия

2.3. Требования к персоналу (пользователю)

Минимальное количество персонала, требуемого для работы программы, должно составлять не менее 1 штатной единицы:

- 1) конечный пользователь – оператор ЭВМ.

Оператор ЭВМ должен:

- 1) иметь образование не ниже среднего (полного) общего;
- 2) обладать практическими навыками работы с пользовательским интерфейсом операционной системы.
- 3) уметь базовые знания по кластерного анализу, по нормализации данных
- 4) уметь редактировать XML документы

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3. ВЫПОЛНЕНИЕ ПРОГРАММЫ

В данном разделе показано, как устанавливать программу и пользоваться ей.

3.1. Установка и запуск программы

Дистрибутив программы «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода», поставляется в виде установщика на компакт-диске. В состав поставки программы входит само приложение, техническая документация и презентация проекта.

Для установки программы необходимо запустить мастер установки, дважды нажав на его иконку. На экране появиться приветственное окно установщика (рис. 1):

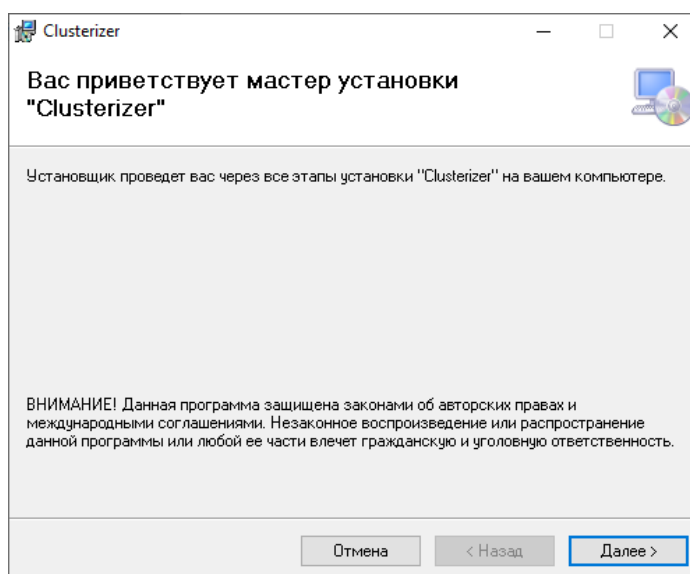


Рисунок 1

Для продолжения установки программы, нужно в появившемся окне нажать на кнопку “Далее”. Затем в новом диалоговом окне (рис. 2) необходимо указать путь размещения приложения.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

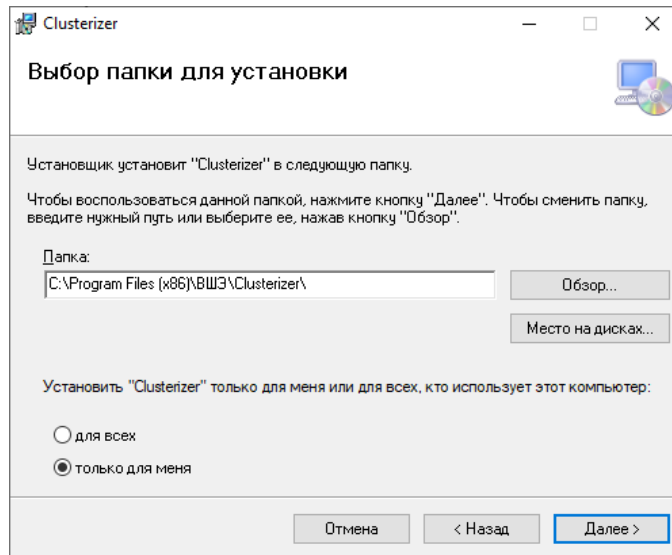


Рисунок 2

Также пользователь может указать, кто имеет право на использование программой, выбрав один из переключателей “для всех” или “только для меня”. Чтобы перейти на следующий шаг установки, нужно нажать на кнопку “Далее”. В новом окне (рис. 3) необходимо подтвердить начало установки программы путем нажатия кнопки “Далее”.

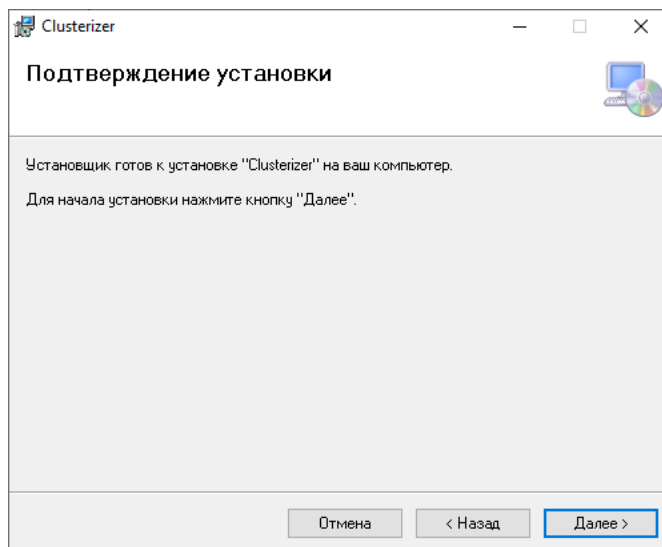


Рисунок 3

После этого начнется установка программы по указанному пути. После завершения установки, появится диалоговое окно (рис. 4), информирующее пользователя об успешности установки программы:

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

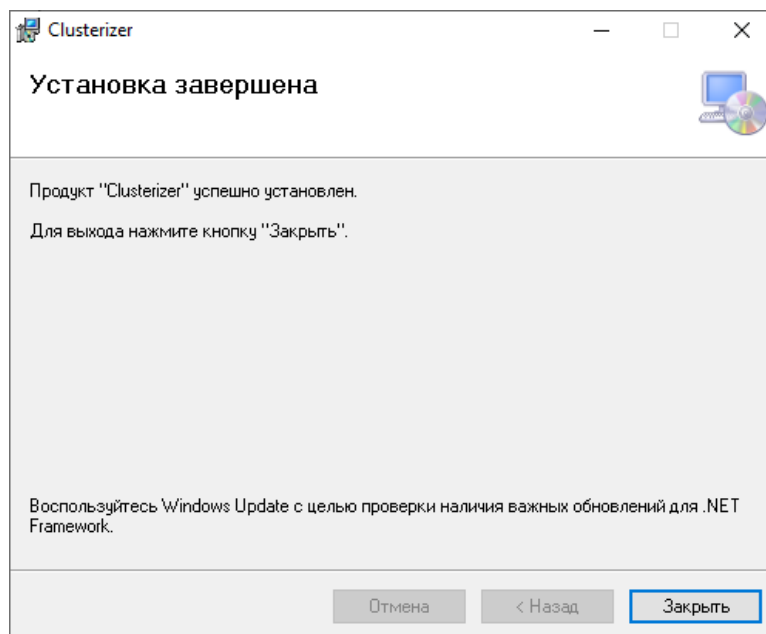


Рисунок 4

Вместе с установкой будут добавлены ярлыки программы на рабочий стол и в меню “Пуск”. Для запуска программы следует нажать на один из ярлыков. При этом может потребоваться предоставить программе права администратора.

3.2. Стартовая страница

При запуске программы открывается стартовая страница (рис. 5):

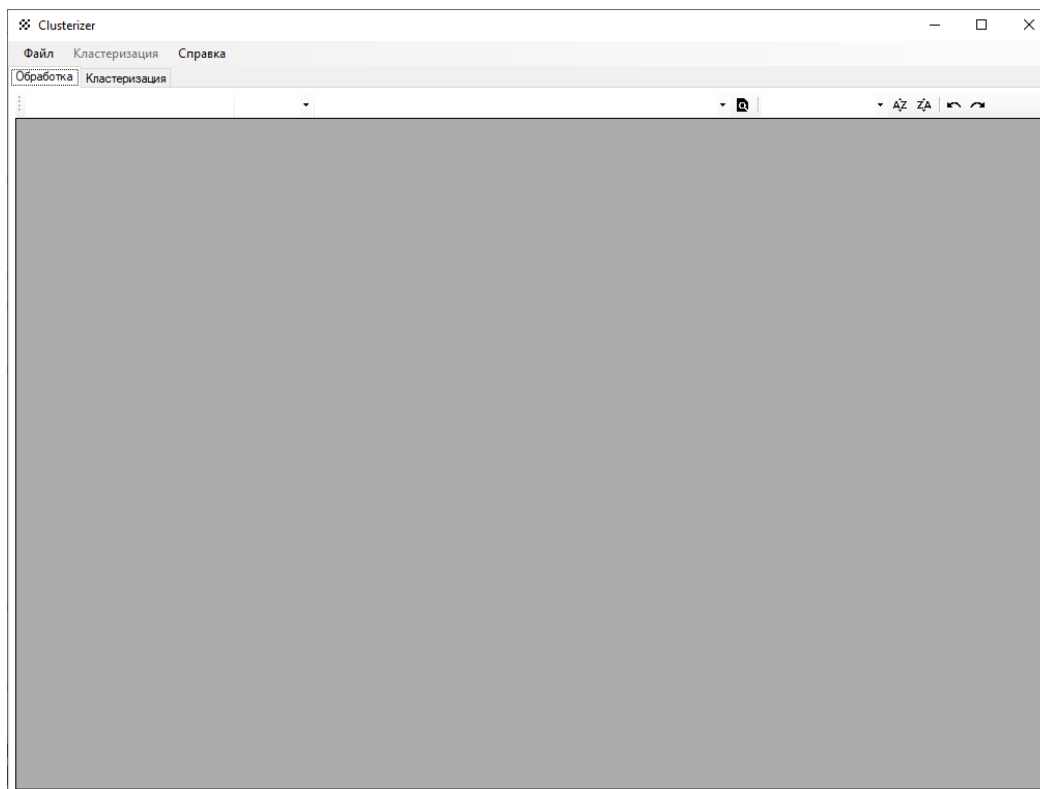


Рисунок 5 – Стартовая страница программы

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3.3. Подготовка программы к работе

Перед открытием программы нужно отредактировать файл dataconfig.xml лежащий в корне программы, в соответствии данных, по которым должна быть выполнена работа.

Также в компакт-диске находится папка “Примеры” с тестовым входным файлом и dataconfig.xml. Для открытия тестового файла скопируйте dataconfig.xml в корень программы.

В качестве входных данных принимаются файлы формата CSV [13] в специальном формате (см. пример в Приложение 4), которая задается через конфигурационный XML файл dataconfig.xml(см. пример в Приложение 3).

В dataconfig.xml содержатся следующие элементы:

- StringHeadings – строковой массив, элементы которого задают описание объекта кластеризации (далее строковые данные). Т. е. элементы, по которым не будет проведена кластеризация. Например, такие характеристики как название объекта, местоположение объекта, тип объекта.
- NumericHeadings – строковой массив, элементы которого задают имена характеристик объекта по которым может быть проведена кластеризация (далее показатели).
- GroupNames – строковой массив, элементы которого задают имена групп по которым разделяются показатели.
- GroupItemsCount – массив натуральных чисел, элементы которого задают количество показателей в каждой группе.

При этом каждый из элементов не может быть пустым и сумма элементов

GroupItemsCount должна равна сумме количества элементов StringHeadings и NumericHeadings. В качестве первого элемента StringHeadings указывается обусловленное название объекта (далее название кластера).

3.4. Работа с входными данными

3.1. Открытие файла

Для открытия входного файла нужно нажать на кнопку “Открыть” из меню “Файл”. После чего появиться окно (рис. 6) для выбора входного файла.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

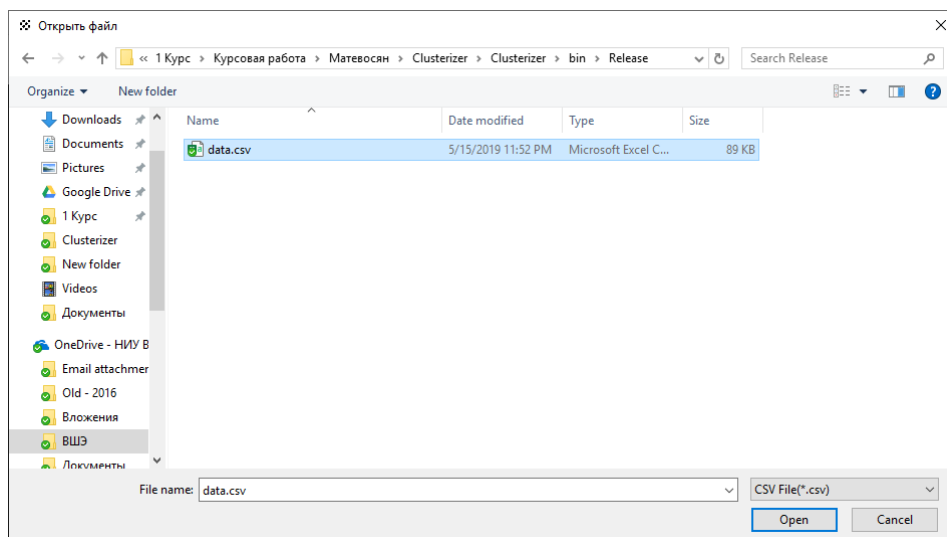


Рисунок 6 – Окно открытия файла

Нужно выбрать нужный файл и нажать на кнопку “Открыть”.

3..2. Заккрытие файла

Для сохранения входного файла нужно нажать на кнопку “Заккрыть” из меню “Файл”.

3..3. Сохранение файла

Для сохранения входного файла нужно нажать на кнопку “Сохранить” из меню “Файл”.

Для сохранения входного файла в другой файл, нужно нажать на кнопку “Сохранить как...”. После чего появится окно (рис. 7) для выбора сохраняемого файла.

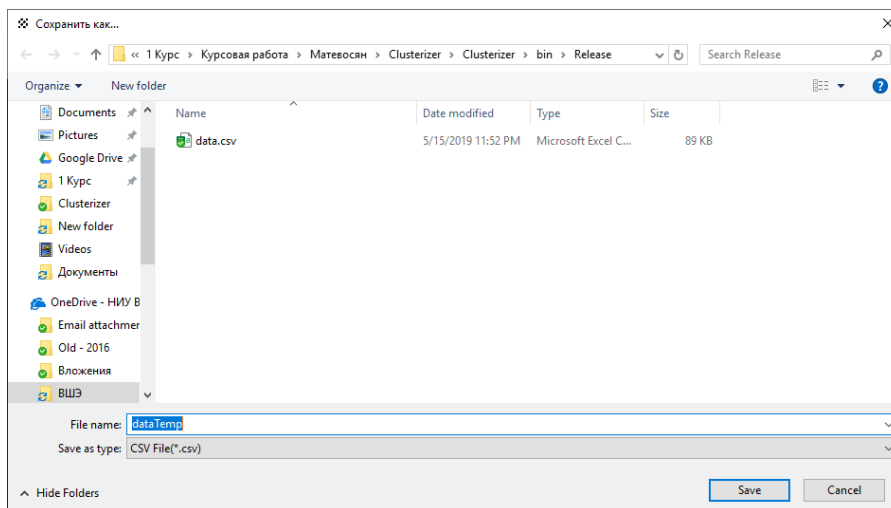


Рисунок 7 – Окно сохранения файла

Нужно выбрать место и имя сохраняемого файла и нажать на кнопку “Сохранить”.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3.5. Работа с таблицей данных

3..1. Добавление новой строки

Для добавления новой строки нужно нажать на кнопку (*) в конце таблицы данных (рис. 8). После чего добавиться новая строка с значениями по умолчанию.



Рисунок 8

3..2. Изменение данных в таблице

Чтобы изменить данные нужно выбрать элемент, значение которого вы хотите изменить и ввести новое значение.

3..3. Удаление строки

Чтобы удалить строку, нужно выбрать ее и нажать на кнопку “Delete” клавиатуры.

3..4. Фильтрация данных

Для фильтрации данных нужно сначала выбрать столбец фильтрации, потом уже условие фильтрации. Потом ввести значение и нажать на кнопку фильтрации из панели инструментов (рис. 9).



Рисунок 9

3..5. Сортировка данных

Для сортировки данных нужно сначала выбрать столбец сортировки, а потом нажать на соответствующую кнопку сортировки из панели инструментов (рис. 10). Может выполняться сортировка по возрастанию и сортировка по убыванию.

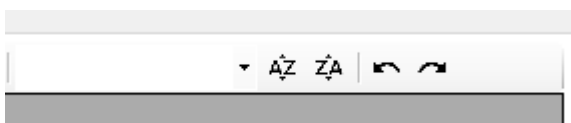


Рисунок 10

3.6. Кластеризация данных

3..1. Выполнение кластеризации данных

Для кластеризации данных нужно нажать на кнопку “Выполнить кластеризацию” из меню “Кластеризация”. Появится окно (рис. 11) для выбора параметров кластеризации.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

Мера расстояния: Евклидово расстояние

Стратегия объединения: Одноточечная связь (расстояния ближайшего соседа)

Метод нормализации: Никакой

Число кластеров: 2

Рассчитать рекомендуемое число кластеров

Выберите показатели по которым нужно выполнить кластеризацию:

- ☒ Образовательная деятельность
- ☒ Научно-исследовательская деятельность
- ☒ Международная деятельность
- ☒ Финансово-экономическая деятельность
- ☒ Инфраструктура
- ☒ Трудоустройство
- ☒ Кадровый состав

Выполнить кластеризацию

Рисунок 11 – Окно выбора параметров кластеризации

По умолчанию выбираются все показатели для кластеризации. Если стоит “птичка” под показателем значит она выбрана. При клике на кнопку “+” перед ним раскроется группа со всеми показателями внутри.

Так же нужно ввести выходное количество кластеров и выбрать другие параметры кластеризации. Для расчета рекомендуемого количества кластеров нужно нажать на кнопку “Рассчитать рекомендуемое число кластеров”.

3..2. Построить дендограмму

Чтобы построить дендограмму нужно нажать на кнопку “Построить дендограмму” из меню “Кластеризация”. Появится окно (рис. 12) с полученной дендограммой.

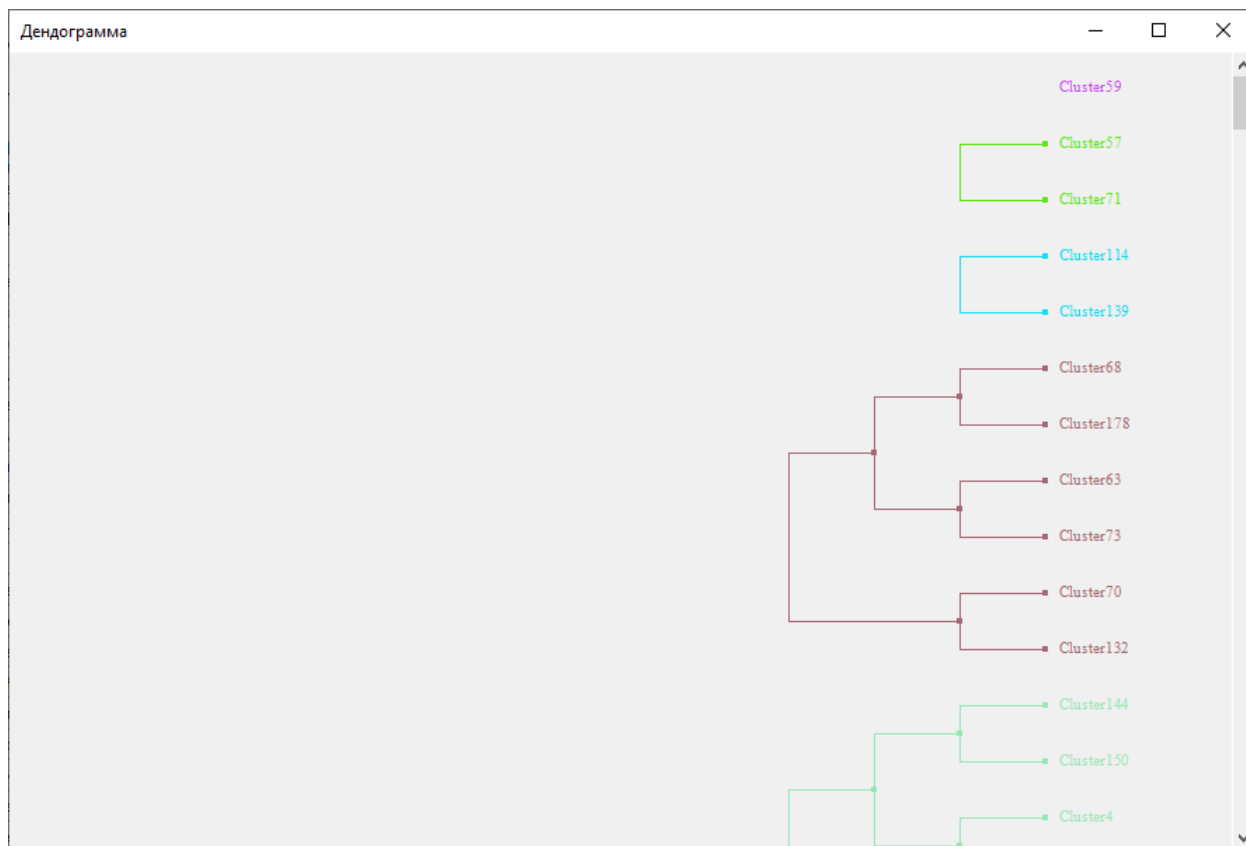


Рисунок 12 – Окно с полученной дендограммой

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3..3. Показать описательную статистику кластеров

Чтобы показать описательную статистику кластеров нужно нажать на кнопку “Показать статистику кластеров” из меню “Кластеризация”. Появится окно (рис. 13) с полученной таблицей.

Статистика кластеров				
	Название кластера	Средний балл ЕГЭ студентов, принятых на обучение по очной форме за счет средств соответствующих бюджетов бюджетной системы РФ	Средний балл ЕГЭ студентов, принятых на обучение по очной форме за счет средств соответствующих бюджетов бюджетной системы РФ, за исключением лиц, поступивших с учетом особых прав и в рамках квоты целевого приема	Средний балл ЕГЭ студентов, принятых на обучение по очной форме с оплатой стоимости затрат на обучение физическими и юридическими лицами
1	Cluster59	88.38	86.66	74.02
2	Cluster288	87.52	84.92	72.27
3	Cluster291	39.44	81.785	73.22
4	Cluster377	84.79333333333333	83.46166666666667	72.76666666666667
5	Cluster378	70.0332692307692	69.6690384615384	60.1894230769231
6	Cluster379	77.785	77.026	64.732
7	Cluster380	67.2675961538462	67.6754807692308	58.3964423076923
8	Cluster381	70.24944444444444	69.97833333333333	59.84333333333333

Рисунок 13 – Окно статистики кластеров

3.7. Экспорт

Для экспорта выходных файлов нужно нажать выбрать тип экспортируемого файла из пункта Экспорт (рис. 14) меню “Файл”.

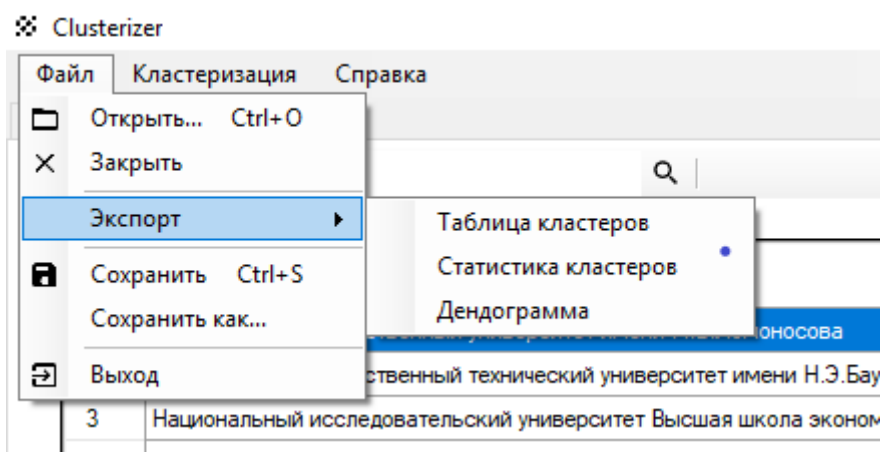


Рисунок 14

После чего появится окно (рис. 15) для выбора сохраняемого файла.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

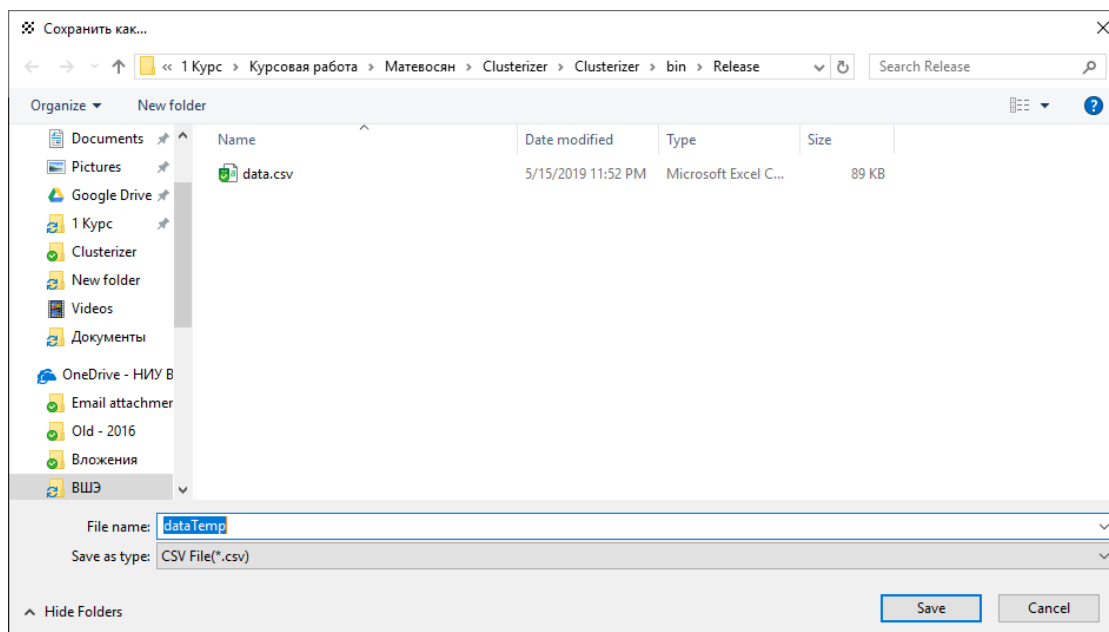


Рисунок 15

Нужно выбрать место и имя сохраняемого файла и нажать на кнопку “Сохранить”.

3.8. О программе

Для просмотра информации о программе нужно нажать на кнопку “О программе” из пункта “Справка” меню. Появится с окно (рис. 16) с информацией о программе.

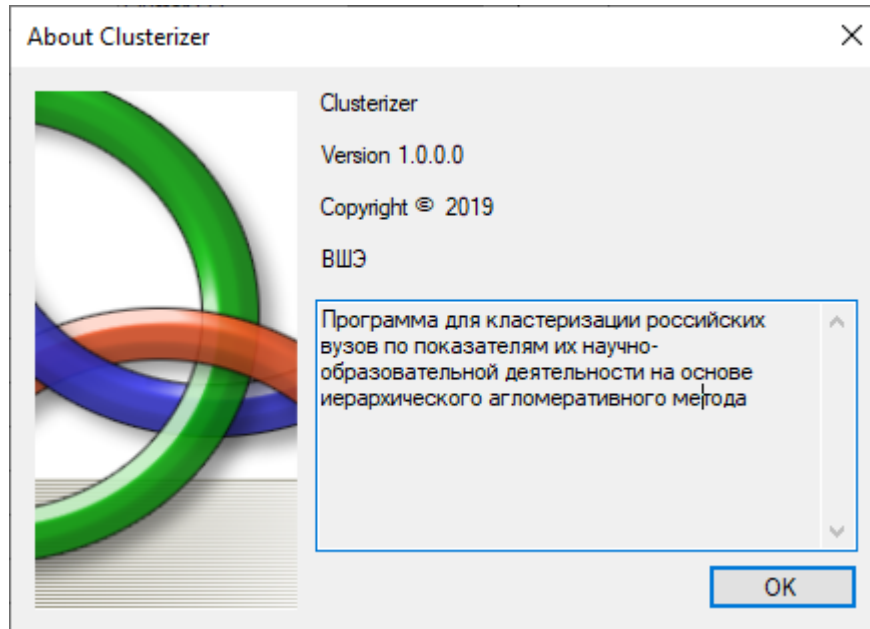


Рисунок 16 – Окно информации о программе

3.9. Завершение работы с программой

Для завершения программы нужно нажать на кнопку “Закрыть” из пункта “Файл” меню.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4. СООБЩЕНИЯ ОПЕРАТОРУ

Для удобства пользователя в программе предусмотрены всплывающие диалоговые окна, отображающие некоторую информацию.

1. Сообщение об ошибке при чтении конфигурационного файла (т.е. конфигурационный файл поврежден) (рис. 17):

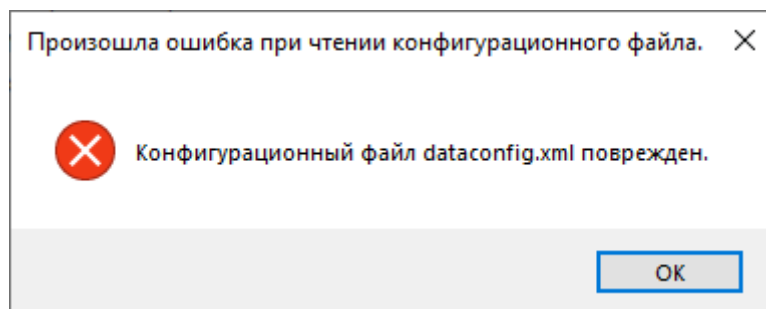


Рисунок 17

2. Сообщение об ошибке при открытии входных данных (т.е. данные не правильного формата, были открыты в другом приложении или нет доступа к файлу) (рис. 18):

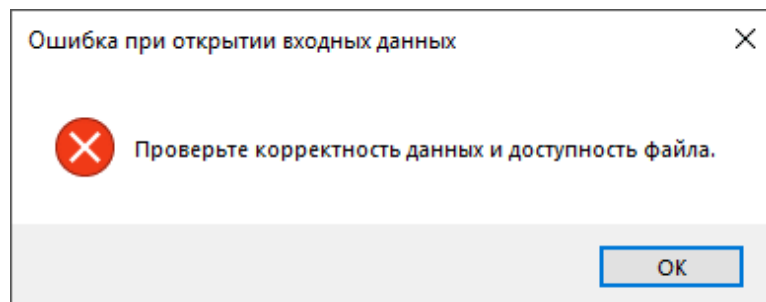


Рисунок 18

3. Сообщение об ошибке редактировании таблицы(т.е. введено некорректное значение в таблице) (рис. 19):

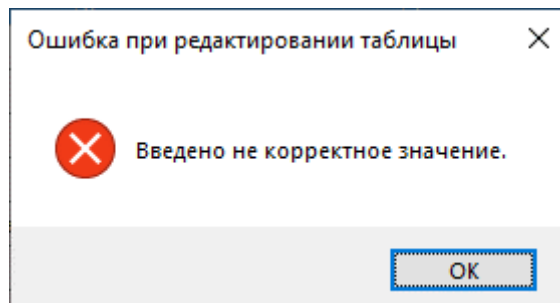


Рисунок 19

4. Сообщение об ошибке при выборе показателей (т.е. пользователь не выбрал ни одного показателя кластеризации) (рис. 20):

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

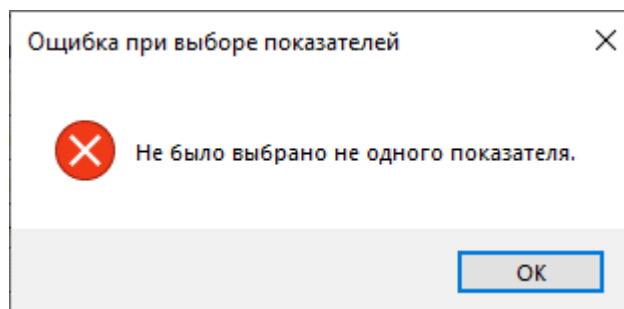


Рисунок 20

5. Сообщение об ошибке при вводе числа кластеров (т.е. пользователь ввел не корректное значение числа кластеров) (рис. 21):

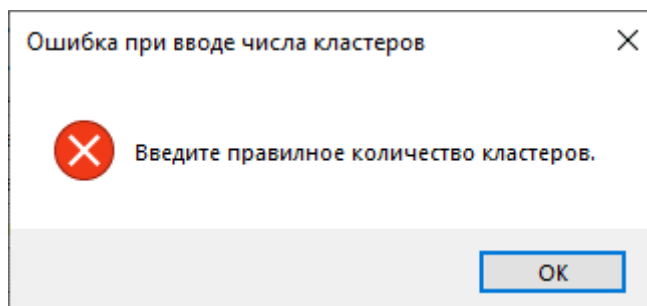


Рисунок 21

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

5. 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 19.505-79 Руководство оператора. 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) Мандель И. Д. Кластерный анализ. — М.: Финансы и статистика, 1988. — 176 с.
- 11) Шрейдер Ю. А. Что такое расстояние? — М.: Физматлит, 1963. — 76 с.
- 12) Жамбю М. Иерархический кластер-анализ и соответствия. — М.: Финансы и статистика, 1988. — 345 с.
- 13) Common Format and MIME Type for Comma-Separated Values (CSV) Files. [Электронный ресурс] / SolidMatrix Technologies, Inc.: <https://tools.ietf.org/html/rfc4180>, свободный(дата обращения: 19.04.2019).
- 14) Windows 7 system requirements – Windows Help. [Электронный ресурс] / Microsoft. Режим доступа: <https://support.microsoft.com/en-us/help/10737/windows-7-system-requirements>, свободный(дата обращения: 19.04.2019).

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ТЕРМИНОЛОГИЯ

Ниже приведен список необходимых терминов для ознакомления [10], [11], [12].

Агломеративный метод – один из методов иерархической кластеризации, в котором создание новых кластеров выполняется путем объединения малочисленных кластеров в более крупные кластеры, таким образом дерево созданным методом имеет направление от листьев к стволу, которая называется деревом кластеров.

Дендрограмма – совокупность древовидных диаграмм дерева кластеров.

Иерархические алгоритм – группа алгоритмов кластеризации, которая упорядочивает данные путем создания иерархии(дерева) вложенных кластеров.

Индекс Цалиньски Харабаша – критерия для оценки обусловленного качества выполненной кластеризации.

Кластер — группа однородных объектов.

Кластеризация (или кластерный анализ) — задача группирования множества объектов так, чтобы объекты, которые принадлежат одной группе были более похожими(однородными), а объекты разных групп должны максимально быть различны. Сама кластеризация не является алгоритмом, а общей задачей для решения.

Матрица различия — матрица в котором хранятся значения расстояния между двумя кластерами.

Мера расстояния — метрика, которая описывает расстояние между двумя объектами.

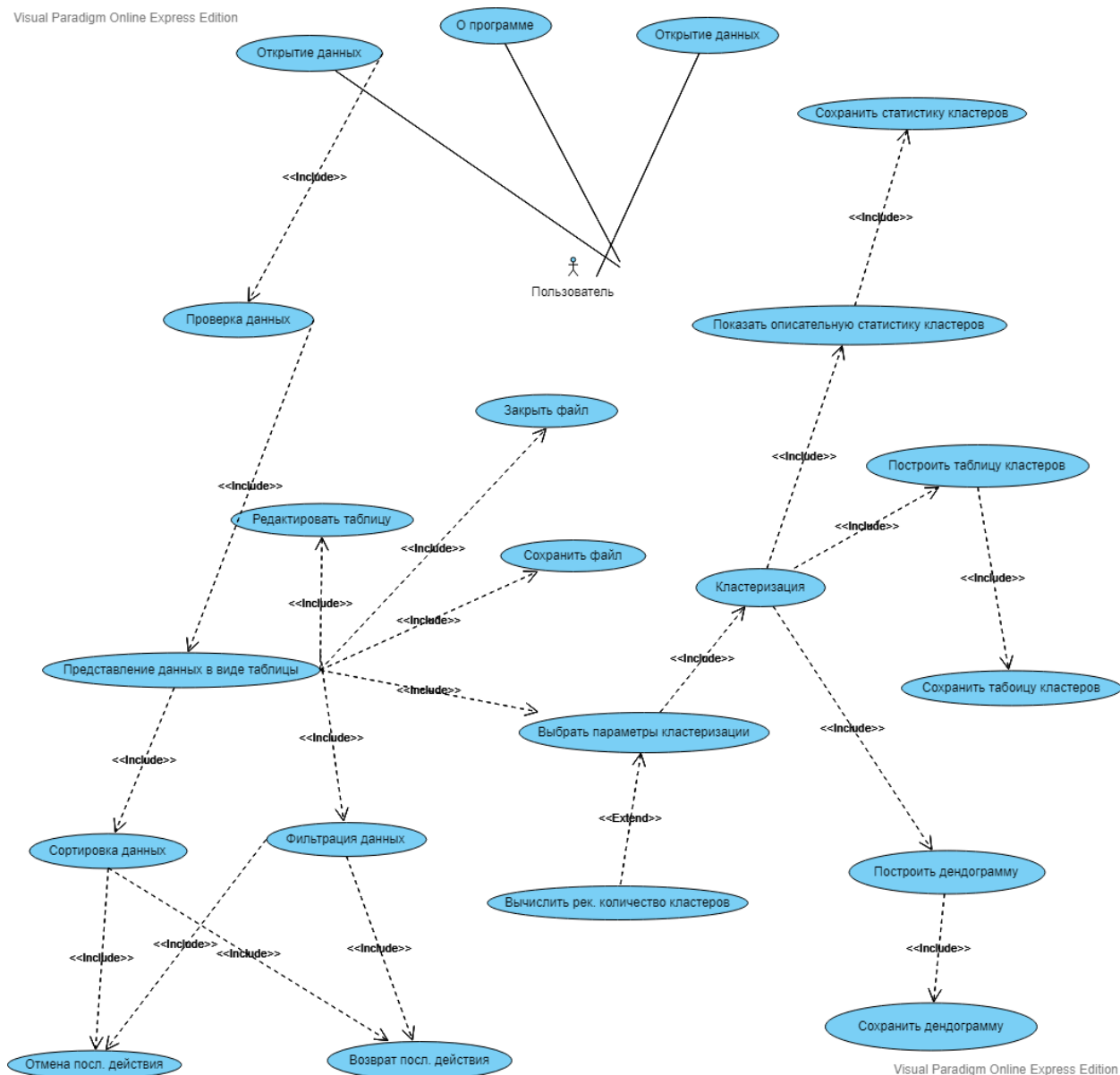
Стратегия объединения — алгоритм объединения двух кластеров.

Сумма квадратического отклонения кластера – сумма квадратов расстояния объектов кластера от его центроида.

Центроид – центр тяжести кластера. Представляет собой кластер, значения точек(характеристик) которой равны среднему значению по каждой точке.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ДИАГРАММА ПРЕЦЕДЕНТОВ



Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИМЕР КОНФИГУРАЦИОННОГО ФАЙЛА dataconfig.xml

```
<?xml version="1.0"?>
<Configuration xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <StringHeadings>
    <string>Название вуза</string>
    <string>Округ</string>
    <string>Субъект РФ</string>
    <string>Город</string>
    <string>Ведомственная принадлежность</string>
    <string>Профиль организации</string>
  </StringHeadings>
  <NumericHeadings>
    <string>Удельный вес выпускников, трудоустроившихся в течение календарного года,
    следующего за годом выпуска, в общей численности выпускников образовательной
    организации обучавшихся по основным образовательным программам высшего
    образования</string>
    <string>Удельный вес НПП, имеющих ученую степень кандидата наук, в общей
    численности НПП</string>
    <string>Удельный вес НПП имеющих ученую степень доктора наук, в общей
    численности НПП</string>
    <string>Удельный вес НПП, имеющих ученую степень кандидата и доктора наук, в
    общей численности НПП образовательной организации (без совместителей и работающих
    по договорам гражданско-правового характера)</string>
    <string>Число НПП, имеющих ученую степень кандидата и доктора наук, в расчете на
    100 студентов</string>
    <string>Доля штатных работников ППС в общей численности ППС</string>
  </NumericHeadings>
  <GroupNames>
    <string>Трудоустройство</string>
    <string>Кадровый состав</string>
  </GroupNames>
  <GroupItemsCount>
    <int>1</int>
    <int>5</int>
  </GroupItemsCount>
</Configuration>
```

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 4**ПРИМЕР ВХОДНОГО ФАЙЛА**

Российский новый университет;1;1;Москва;1;1;0;50.15;21.49;67.78;1.25;79.55
 Адыгейский государственный университет;6;60;Майкоп;5;1;0;74.19;15.9;90.93;6.77;87.56
 Алтайский государственный медицинский университет Министерства здравоохранения
 Российской Федерации;3;9;Барнаул;9;4;0;57.59;22.1;78.7;9.59;80.08
 Алтайский государственный технический университет им. И.И.
 Ползунова;3;9;Барнаул;5;1;75;59.64;10.95;70.39;4.55;93.35
 Алтайский государственный университет;3;9;Барнаул;5;1;0;64.24;16.72;82.11;4.81;87.11
 Астраханский государственный медицинский университет Министерства
 здравоохранения Российской Федерации;6;57;Астрахань;9;4;65;51.65;19.65;70.47;8.77;75.5
 Астраханский государственный технический
 университет;6;57;Астрахань;20;1;70;59.07;14.14;74.07;4.05;76.52
 Астраханский государственный
 университет;6;57;Астрахань;5;1;70;60.14;15.3;75.32;4.64;77.32
 Балтийский федеральный университет имени Иммануила
 Канта;5;49;Калининград;5;1;70;48.67;14.96;66.87;5.77;82.86
 Башкирский государственный медицинский университет Министерства здравоохранения
 Российской Федерации;9;77;Уфа;9;4;85;64.9;26.98;91.37;10.35;76.09
 Башкирский государственный университет;9;77;Уфа;5;1;75;63.63;23.03;84.11;5.07;85.7
 Белгородский государственный национальный исследовательский
 университет;1;30;Белгород;5;1;0;60.21;16.07;78.52;5.44;83.07
 Белгородский государственный технологический университет им. В.Г.
 Шухова;1;30;Белгород;5;1;75;58.89;17.07;70.37;3.93;87.32
 Благовещенский государственный педагогический
 университет;4;21;Благовещенск;5;1;75;70.73;9.24;80.43;4.07;92.93
 Брянский государственный аграрный университет;1;31;село
 Кокино;6;3;65;63.4;19.79;82.83;3.35;96.94
 Владивостокский государственный университет экономики и
 сервиса;4;25;Владивосток;5;1;65;62.57;9.34;67.72;3.15;83.53
 Владимирский государственный университет имени Александра Григорьевича и Николая
 Григорьевича Столетовых;1;32;Владимир;5;1;80;58.27;13.28;72.48;3.84;84.04
 Волгоградский государственный медицинский университет Министерства
 здравоохранения Российской
 Федерации;6;58;Волгоград;9;4;80;53.47;15.42;71.44;11.69;72.4

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 5**ПРИМЕР ВЫХОДНОГО ФАЙЛА ТАБЛИЦЫ КЛАСТЕРОВ**

Адыгейский государственный университет;Cluster1;Cluster26
 Благовещенский государственный педагогический университет;Cluster13;Cluster26
 Российский новый университет;Cluster0;Cluster29
 Алтайский государственный медицинский университет Министерства здравоохранения
 Российской Федерации;Cluster2;Cluster29
 Алтайский государственный университет;Cluster4;Cluster29
 Белгородский государственный национальный исследовательский
 университет;Cluster11;Cluster29
 Астраханский государственный медицинский университет Министерства
 здравоохранения Российской Федерации;Cluster5;Cluster32
 Балтийский федеральный университет имени Иммануила Канта;Cluster8;Cluster32
 Астраханский государственный технический университет;Cluster6;Cluster32
 Астраханский государственный университет;Cluster7;Cluster32
 Алтайский государственный технический университет им. И.И.
 Ползунова;Cluster3;Cluster32
 Владивостокский государственный университет экономики и сервиса;Cluster15;Cluster32
 Белгородский государственный технологический университет им. В.Г.
 Шухова;Cluster12;Cluster32
 Владимирский государственный университет имени Александра Григорьевича и Николая
 Григорьевича Столетовых;Cluster16;Cluster32
 Башкирский государственный университет;Cluster10;Cluster32
 Брянский государственный аграрный университет;Cluster14;Cluster32
 Башкирский государственный медицинский университет Министерства здравоохранения
 Российской Федерации;Cluster9;Cluster32
 Волгоградский государственный медицинский университет Министерства
 здравоохранения Российской Федерации;Cluster17;Cluster32

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 6

ПРИМЕР ВЫХОДНОГО ФАЙЛА ОПИСАТЕЛЬНОЙ СТАТИСТИКИ КЛАСТЕРОВ

Cluster26;37.5;72.46;12.57;85.68;5.42;90.245;2

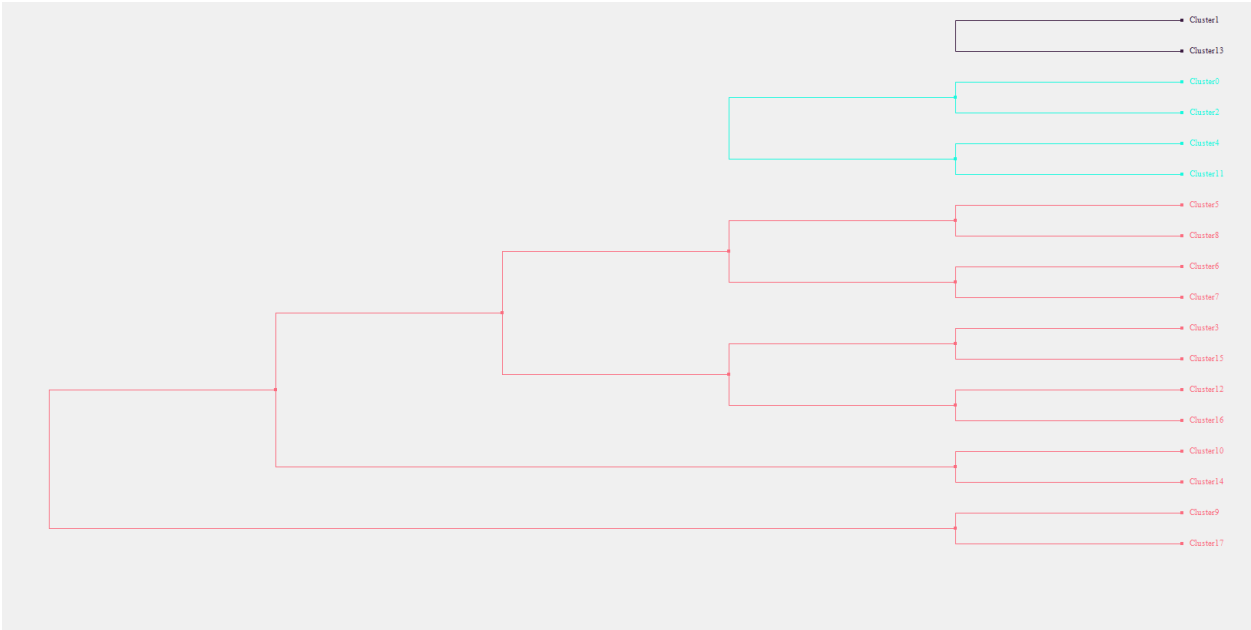
Cluster29;0;58.0475;19.095;76.7775;5.2725;82.4525;4

Cluster32;72.9166666666667;58.6916666666667;16.6591666666667;74.7866666666667;5.763
33333333333;82.6308333333333;12

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 7

ПРИМЕР ВЫХОДНОГО ФАЙЛА КАРТИНКИ ДЕНДОГРАММЫ



Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

[illegible]

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13 34				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата