


ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Департамент программной инженерии

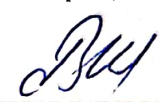
СОГЛАСОВАНО

Научный руководитель,
старший преподаватель департамента
программной инженерии факультета
компьютерных наук


«22» мал А.В. Меликян
2019 г.

УТВЕРЖДАЮ

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


«22» мал В.В. Шилов
2019 г.

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

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

Техническое задание

ЛИСТ УТВЕРЖДЕНИЯ

RU.17701729.04.13-01 ТЗ 01-1-ЛУ

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

Москва 2019

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

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

Техническое задание

RU.17701729.04.13-01 ТЗ 01-1

Листов 27

<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

Техническое задание – это основной документ, оговаривающий набор требований и порядок создания программного продукта, в соответствии с которым производится разработка программы, ее тестирование и приемка.

Настоящее Техническое задание на разработку «Программы для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода» содержит следующие разделы: «Введение», «Основание для разработки», «Назначение разработки», «Требования к программе», «Требования к программным документам», «Технико-экономические показатели», «Стадии и этапы разработки», «Порядок контроля и приемки» и приложения [7].

В разделе «Введение» указано наименование и краткая характеристика области применения «Программы для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода».

В разделе «Основания для разработки» указан документ на основании, которого ведется разработка и наименование темы разработки.

В разделе «Назначение разработки» указано функциональное и эксплуатационное назначение программного продукта.

Раздел «Требования к программе» содержит основные требования к функциональным характеристикам, к надежности, к условиям эксплуатации, к составу и параметрам технических средств, к информационной и программной совместимости, к маркировке и упаковке, к транспортировке и хранению, а также специальные требования.

Раздел «Требования к программным документам» содержит предварительный состав программной документации и специальные требования к ней.

Раздел «Технико-экономические показатели» содержит ориентировочную экономическую эффективность, предполагаемую годовую потребность, экономические преимущества разработки «Программы для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода».

Раздел «Стадии и этапы разработки» содержит стадии разработки, этапы и содержание работ.

В разделе «Порядок контроля и приемки» указаны общие требования к приемке работы.

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 19.201-78 Техническое задание. Требования к содержанию и оформлению [7].

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

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

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

СОДЕРЖАНИЕ

1.	ВВЕДЕНИЕ	5
1.1.	Наименование программы	5
1.2.	Краткая характеристика области применения программы	5
2.	ОСНОВАНИЕ ДЛЯ РАЗРАБОТКИ.....	6
2.1.	Документы, на основании которых ведется разработка.....	6
2.2.	Наименование темы разработки	6
3.	НАЗНАЧЕНИЕ РАЗРАБОТКИ	7
3.1.	Функциональное назначение.....	7
3.2.	Эксплуатационное назначение	7
4.	ТРЕБОВАНИЯ К ПРОГРАММЕ	8
4.1.	Требования к функциональным характеристикам	8
4.1.1.	Требования к составу выполняемых функций	8
4.1.2.	Требования к организации входных данных.....	8
4.1.3.	Требования к организации выходных данных	9
4.1.4.	Требования к временным характеристикам	9
4.2.	Требования к интерфейсу	9
4.3.	Требования к надежности	9
4.3.1.	Требования к обеспечению надежного (устойчивого) функционирования программы	9
4.3.2.	Время восстановления после отказа	9
4.3.3.	Отказы из-за некорректных действий оператора	10
4.4.	Условия эксплуатации	10
4.4.1.	Климатические условия эксплуатации	10
4.4.2.	Требования к видам обслуживания.....	10
4.4.3.	Требования к численности и квалификации персонала	10
4.5.	Требования к составу и параметрам технических средств.....	11
4.6.	Требования к информационной и программной совместимости	11
4.6.1.	Требования к информационным структурам и методам решения	11
4.6.2.	Требования к программным средствам, используемым программой.	11
4.6.3.	Требования к исходным кодам и языкам программирования.....	11
4.6.4.	Требования к защите информации и программы.....	11
4.7.	Требования к маркировке и упаковке.....	11
4.8.	Требования к транспортировке и хранению	11
4.8.1.	Требования к хранению и транспортировке компакт-дисков (CD)	11
4.8.2.	Требования к хранению и транспортировке программных документов, предоставляемых в печатном виде.	12
4.9.	Специальные требования.....	12
5.	ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ	13
5.1.	Состав программной документации	13
5.2.	Специальные требования к программной документации.....	13
6.	ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ	14
6.1.	Предполагаемая потребность	14

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

6.2.	Ориентировочная экономическая эффективность	14
6.3.	Экономические преимущества разработки по сравнению с отечественными и зарубежными аналогами	14
7.	СТАДИИ И ЭТАПЫ РАЗРАБОТКИ	15
7.1.	Стадии разработки.....	15
7.2.	Сроки разработки и исполнители	16
8.	ПОРЯДОК КОНТРОЛЯ И ПРИЁМКИ.....	17
9.	СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ.....	18
	ПРИЛОЖЕНИЕ 1 - ТЕРМИНАЛОГИЯ	19
	ПРИЛОЖЕНИЕ 2 - ДИАГРАММА ПРЕЦЕДЕНТОВ	20
	ПРИЛОЖЕНИЕ 3 - ПРИМЕР КОНФИГУРАЦИОННОГО ФАЙЛА DATACONFIG.XML	21
	ПРИЛОЖЕНИЕ 4 - ПРИМЕР ВХОДНОГО ФАЙЛА	22
	ПРИЛОЖЕНИЕ 5 - ПРИМЕР ВЫХОДНОГО ФАЙЛА ТАБЛИЦЫ КЛАСТЕРОВ	23
	ПРИЛОЖЕНИЕ 6 - ПРИМЕР ВЫХОДНОГО ФАЙЛА ОПИСАТЕЛЬНОЙ СТАТИСТИКИ КЛАСТЕРОВ.....	24
	ПРИЛОЖЕНИЕ 7 - ПРИМЕР ВЫХОДНОГО ФАЙЛА КАРТИНКИ ДЕНДОГРАММЫ ...	25
	ПРИМЕР ВЫХОДНОГО ФАЙЛА КАРТИНКИ ДЕНДОГРАММЫ КЛАССОВ	25
	ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ.....	26

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

1. ВВЕДЕНИЕ

1.1. Наименование программы

Наименование программы: «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода» («Program for Clustering Russian Universities on their Educational and Research Indicators on the Basis of Agglomerative Hierarchical Method»).

Краткое наименование программы: («Clusterizer»)

1.2. Краткая характеристика области применения программы

«Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода» - программа, которая является инструментом для кластерного анализа данных.

Классы или концептуально смысловые группы являются объектами, которые разделяют общие характеристики, что играет важную роль в том, как люди понимают и описывают окружающий мир. С самого рождения мы обучены тому, чтобы разделять объекты на некоторые группы. В контексте понимания данных кластеры потенциальные классы, а кластерный анализ является техникой для автоматического нахождения этих классов.

Кластерный анализ и ее различные реализации могут быть использованы в разных областях для решения задач, таких как:

1. Классификация различных биологических видов, элементов
2. В разбиении результатов по запросу в поисковых системах (Google, Yahoo, Yandex...)
3. В различных исследованиях для выявления закономерностей
4. В нахождении вариации различных болезней в медицине
5. В бизнес среде, для обработки большого объема данных

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

2. ОСНОВАНИЕ ДЛЯ РАЗРАБОТКИ

2.1. Документы, на основании которых ведется разработка

Разработка «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода» ведется на основании Приказа № 2.3-02/0812-01 от 08.12.16 «Об утверждении тем, руководителей курсовых работ студентов образовательной программы Программная инженерия факультета компьютерных наук».

Разработка выполняется в рамках темы курсовой работы в соответствии с учебным планом подготовки бакалавров Национального исследовательского университета «Высшая школа экономики» по направлению 09.03.04 «Программная инженерия».

2.2. Наименование темы разработки

Наименование темы разработки – «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода».

Условное обозначение темы разработки – «Clusterizer».

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

3. НАЗНАЧЕНИЕ РАЗРАБОТКИ

3.1. 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
- Вывод описательной статистики по полученным кластерам

3.2. 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.

Наличие Интернета для работы программы не требуется.

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

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4. ТРЕБОВАНИЯ К ПРОГРАММЕ

4.1. Требования к функциональным характеристикам

4.1.1. Требования к составу выполняемых 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
- 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

В качестве удобства вышеперечисленные функции приведены в диаграмме прецедентов (см. Приложение 2)

4.1.2. Требования к организации входных данных

В качестве входных данных принимаются файлы формата CSV [18] 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 – строковый 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 – строковый массив, элементы которого задают имена характеристик объекта по которым может быть проведена кластеризация (далее показатели).
- GroupNames – строковый массив, элементы которого задают имена групп по которым разделяются показатели.
- GroupItemsCount – массив натуральных чисел, элементы которого задают количество показателей в каждой группе.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

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).

4.1.3. Требования к организации выходных данных

В качестве выходных данных программы выступают:

- Открытый программой входной файл в формате заданном выше
- Таблица кластеров программы, сохраняемая в формате CSV (см. Приложение 5), которая состоит из трех столбцов: название кластера, изначальный кластер и выходной кластер.
- Таблица описательной статистик сохраняемая в формате CSV (см. Приложение 6), показывающая среднее значение показателей в выходном кластере, а также количество элементов в кластере. Столбцами таблицы являются показатели и количество кластеров.
- Картинка дендограммы сохраняемая в формате PNG (см. Приложение 7)

4.1.4. Требования к временным характеристикам

Необходимо оценить временную сложность разных алгоритмов таких как AGNES [19], нормализации данных, разбиения множества на равные подмножества по среднеквадратическому среднему отклонение.

4.2. Требования к интерфейсу

Программа должна иметь удобный и простой в использованный оконный Windows-интерфейс.

Добавление новой строки, удаление и изменение строки происходит в самой таблице.

В программе должны быть использованы следующие элементы управления: меню, панель инструментов.

Для выбора параметров кластеризации должно появиться отдельное окно.

4.3. Требования к надежности

4.3.1. Требования к обеспечению надежного (устойчивого) функционирования программы

Для устойчивой работы программы необходимо соблюдать ряд организационно-технических мер:

- 1) обеспечить бесперебойное питание технических устройств;
- 2) обеспечить высокую защиту технических устройств для работы программы от воздействия шпионских программ, троянских программ, программ-шутков и других видов вредоносного программного обеспечения;
- 3) обеспечить регулярную проверку оборудования и программного обеспечения на наличие сбоев и неполадок;
- 4) обеспечить использование лицензионного программного обеспечения.

4.3.2. Время восстановления после отказа

Если отказ был вызван какими-либо внешними факторами, например, сбоем электропитания, и при этом не произошел непоправимый сбой операционной системы, то

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

время восстановления не должно превышать времени, требующегося на перезагрузку операционной системы и запуск программы.

Если отказ был вызван неисправностью технических средств или непоправимым сбоем операционной системы, то время восстановления не должно превышать времени, необходимого для устранения неисправностей технических и программных средств.

4.3.3. Отказы из-за некорректных действий оператора

Отказ программы возможен из-за некорректных действий пользователя при использовании операционной системой. Для предотвращения случаев отказа программы по причине сбоев при использовании операционной системой следует провести предварительный инструктаж конечного пользователя и обеспечить работу конечного пользователя без предоставления ему прав администратора.

Также при повреждении или неправильно заданного внутреннего конфигурационного файла, программа покажет окно об ошибке и закроется сама.

4.4. Условия эксплуатации

4.4.1. Климатические условия эксплуатации

Климатические условия эксплуатации, при которых должны обеспечиваться заданные характеристики, должны удовлетворять требованиям, предъявляемым к персональным компьютерам и компакт-дискам (CD) в части условий их эксплуатации.

Персональный компьютер предназначен для работы в закрытом отапливаемом помещении со стабильными климатическими условиями категории 4.1 согласно ГОСТ 15150-69 [11].

Для предотвращения повреждений компакт-диска (CD) и сохранения на нем записанной информации необходимо поддерживать следующие климатические условия [12]:

- 1) влажность от 20% до 70%;
- 2) температура от 5°C до 30°C;
- 3) атмосферное давление — от 84 до 106,7 кПа (от 630 до 800 мм рт. ст.).

4.4.2. Требования к видам обслуживания

На персональном компьютере, где производится эксплуатация программы необходимо обеспечить регулярные проверки оборудования и программного обеспечения на наличие сбоев и неполадок. Обеспечить защиту персонального компьютера от воздействия шпионских программ, программ-шуток, троянских программ и других видов вирусов.

4.4.3. Требования к численности и квалификации персонала

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 кластерному анализу, по нормализации данных
- 4) уметь редактировать XML документы

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

4.5. Требования к составу и параметрам технических средств

Для надёжной и бесперебойной работы программы требуется следующий состав технических средств [20]:

- 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

4.6. Требования к информационной и программной совместимости

4.6.1. Требования к информационным структурам и методам решения

Требования к методам решения не предъявляются.

4.6.2. Требования к программным средствам, используемым программой.

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

4.6.3. Требования к исходным кодам и языкам программирования

Программа должна быть написана на языке программирования C# 7.0. В качестве интегрированной среды разработки программы должна быть использована среда Microsoft Visual Studio 2017.

4.6.4. Требования к защите информации и программы

Требования к защите информации и программы не предъявляются.

4.7. Требования к маркировке и упаковке

Программа поставляется в виде программного изделия на внешнем носителе информации – компакт диске (CD), на котором должны содержаться программная документация, приложение (исполняемые файлы, два примера задачи и прочие необходимые для работы программы файлы) и презентация проекта.

Программное изделие должно иметь маркировку с обозначением наименования изделия, темы разработки, фамилии, имени и отчества исполнителя и руководителя разработки, учебной группы и года выпуска изделия.

4.8. Требования к транспортировке и хранению

4.8.1. Требования к хранению и транспортировке компакт-дисков (CD)

Программа поставляется заказчику на внешнем носителе информации – компакт-диске (CD). Документация к программе передается как на компакт-диске вместе с программой, так и в печатном виде.

Требования к транспортировке и хранению компакт-дисков с программным обеспечением являются стандартными и должны соответствовать требованиям ГОСТ Р 7.02-2006 [11]:

- 1) В помещении для хранения компакт-дисков допустимы температура воздуха от 10°C до 20°C и относительная влажность воздуха от 20% до 65%.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

- 2) Максимальные суточные колебания температуры 2°C и относительной влажности воздуха - 5 %.
- 3) Компакт-диски хранят и используют на расстоянии не менее 0,5 м от источников тепла и влаги.
- 4) Компакт-диски хранят в темноте или при освещении рассеянным светом, не содержащим ультрафиолетовое излучение.
- 5) Транспортировка производится вертикально в специальных маркированных контейнерах из безопасных материалов с применением мер по предотвращению ударов контейнеров, перемещению и вибрации компакт-дисков внутри контейнеров, а также проникновения влаги, вредных газов, пыли, солнечных лучей и образованию конденсата внутри контейнеров.
- 6) Очистка компакт-диска производится путем протирания чистым мягким хлопчатобумажным тампоном без длинного ворса, пропитанным водой, этиловым (ГОСТ 18300-87 [12]) или изопропиловым (ГОСТ 9805-84 [13]) спиртом.

4.8.2. Требования к хранению и транспортировке программных документов, предоставляемых в печатном виде.

Требования к транспортировке и хранению программных документов являются стандартными и должны соответствовать общим требованиям хранения и транспортировки печатной продукции:

- 1) В помещении для хранения печатной продукции допустимы температура воздуха от 10°C до 30°C и относительная влажность воздуха от 30% до 60%.
- 2) Документацию хранят и используют на расстоянии не менее 0.5 от источников тепла и влаги. Не допускается хранение печатной продукции в помещениях, где находятся агрессивные агенты – растворители, спирт, бензин.
- 3) Не допускается попадание на документацию агрессивных агентов.
- 4) Транспортировка производится в специальных контейнерах с применением мер по предотвращению деформации документов внутри контейнеров, а также проникновения влаги, вредных газов, пыли, солнечных лучей и образованию конденсата внутри контейнеров.
- 5) Программные документы, предоставляемые в печатном виде должны соответствовать общим правилам учета и хранения программных документов, предусмотренных стандартами Единой системы программной документации и соответствовать требованиям ГОСТ 19.602-78 [14].

4.9. Специальные требования

Специальные требования к данной программе не предъявляются.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

5. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

5.1. Состав программной документации

- «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода». Техническое задание (ГОСТ 19.201-78).
- «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода». Программа и методика испытаний (ГОСТ 19.301-78).
- «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода». Текст программы (ГОСТ 19.401-78).
- «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода». Пояснительная записка (ГОСТ 19.404-79).
- «Программа для кластеризации российских вузов по показателям их научно-образовательной деятельности на основе иерархического агломеративного метода». Руководство оператора (ГОСТ 19.505-79).

5.2. Специальные требования к программной документации

Документы к программе должны быть выполнены в соответствии с ГОСТ 19.106-78 и ГОСТами к каждому виду документа (см. п. 5.1.);

Пояснительная записка должна быть загружена в систему Антиплагиат через LMS «НИУ ВШЭ». Лист, подтверждающий загрузку пояснительной записки, сдается в учебный офис вместе со всеми материалами не позже, чем за день до защиты курсовой работы.;

Вся документация также воспроизводится в печатном виде, она должна быть подписана академическим руководителем образовательной программы 09.03.04 «Программная инженерия», руководителем разработки и исполнителем перед сдачей курсовой работы в учебный офис не позже одного дня до защиты;

Документация и программа также сдается в электронном виде в формате .pdf или .docx. в архиве формата .zip или .rar;

За один день до защиты комиссии все материалы курсового проекта:

- техническая документация,
- программный проект,
- исполняемый файл,
- отзыв руководителя

должны быть загружены одним или несколькими архивами в проект дисциплины «Курсовой проект 2018-2019» в личном кабинете в информационной образовательной среде LMS (Learning Management System) НИУ ВШЭ.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

6. ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

6.1. Предполагаемая потребность

Кластерный анализ является одним из востребованных на сегодняшний день направлений для статической обработки данных и машинного обучения. Его могут использовать те, кому нужно будет кластеризовать большой объем данных.

6.2. Ориентировочная экономическая эффективность

В рамках данной работы расчет экономической эффективности не предусмотрен

6.3. Экономические преимущества разработки по сравнению с отечественными и зарубежными аналогами

Существует много алгоритмов кластеризации, и каждый из них по-своему уникален, нет хорошего или плохого. Чаще всего используются готовые библиотеки кластеризации, которые настраиваются под конкретные данные. Так же для кластеризации используют платную программу IBM SPSS.

Данная программа может кластеризовать любые данные, соответствующие заданной конфигурации. Она легкая в использовании, не потребляет много ресурсов и распространяется бесплатно.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

7. СТАДИИ И ЭТАПЫ РАЗРАБОТКИ

7.1. Стадии разработки

I. Техническое задание

1. Обоснование необходимости разработки
 - Постановка задачи;
 - Сбор исходных материалов
2. Научно-исследовательские работы;
 - Определение структуры входных и выходных данных.
 - Предварительный выбор методов решения задач.
 - Обоснование целесообразности применения ранее разработанных программ.
 - Определение требований к техническим средствам.
 - Обоснование принципиальной возможности решения поставленной задачи
3. Разработка и утверждение технического задания
 - Определение требований к программе;
 - Определение стадий, этапов и сроков разработки программы и документации на неё;
 - Выбор языков программирования.
 - Определение необходимости проведения научно-исследовательских работ на последующих стадиях.
 - Согласование и утверждение технического задания.

II. Технический проект

1. Разработка технического проекта
 - Уточнение структуры входных и выходных данных.
 - Разработка алгоритмов и методов решения задачи и подзадач;
 - Определение формы представления входных и выходных данных
 - Разработка структуры программы.
2. Утверждение технического проекта
 - Разработка пояснительной записки (ГОСТ 19.404-79);
 - Согласование и утверждение технического проекта.

III. Рабочий проект

1. Разработка программы
 - Программирование и отладка программы.
2. Разработка программной документации
 - Разработка программной документации в соответствии с требованиями ГОСТ 19 ЕСПД (Единой системы программной документации).
3. Испытания программы
 - разработка, согласование и утверждение программы и методики испытаний;
 - проведение испытаний программы в соответствии с утверждённой программой и методикой;
 - корректировка программы и программной документации по результатам испытаний

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

IV. Внедрение

1. Подготовка и передача программы

- утверждение даты защиты программного продукта;
- подготовка программы и программной документации для презентации и защиты;
- представление разработанного программного продукта руководителю и получение отзыва;
- загрузка Пояснительной записки в систему Антиплагиат через ЛМС НИУ ВШЭ;
- загрузка материалов курсового проекта (курсовой работы) в ЛМС, проект дисциплины «Курсовой проект 2018-2019» (п. 5.2);
- передача программы и сопутствующей программной документации в учебный офис;
- Защита программного продукта (курсового проекта) комиссии.

7.2. Сроки разработки и исполнители

Разработка должна закончиться к 23 мая 2019 года.

Исполнитель: Матевосян Армен Арсенович, студент группы БПИ181 факультета компьютерных наук НИУ ВШЭ.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

8. ПОРЯДОК КОНТРОЛЯ И ПРИЁМКИ

Проверка программного продукта, в том числе и на соответствие техническому заданию, осуществляется исполнителем вместе с заказчиком согласно «Программе и методике испытаний», а также пункту 5.2.

Защита выполненного проекта осуществляется комиссии, состоящей из преподавателей департамента программной инженерии, в утверждённые приказом декана ФКН сроки.

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

9. 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 15150-69 Машины, приборы и другие технические изделия. Исполнения для различных климатических районов. Категории, условия эксплуатации, хранения и транспортирования в части воздействия климатических факторов внешней среды. – М.: Изд-во стандартов, 1997.
- 11) ГОСТ Р 7.02-2006 Консервация документов на компакт-дисках. Общие требования. – М.: ИПК Издательство стандартов, 2006.
- 12) ГОСТ 18300-87 Спирт этиловый ректификованный технический. Технические условия. – М.: ИПК Издательство стандартов, 1997.
- 13) ГОСТ 9805-84. Спирт изопропиловый. Технические условия. – М.: ИПК Издательство стандартов, 1984.
- 14) ГОСТ 19.602-78 Правила дублирования, учета и хранения программных документов, выполненных печатным способом. //Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
- 15) Жамбю М. Иерархический кластер-анализ и соответствия. — М.: Финансы и статистика, 1988. — 345 с.
- 16) Мандель И. Д. Кластерный анализ. — М.: Финансы и статистика, 1988. — 176 с.
- 17) Шрейдер Ю. А. Что такое расстояние? — М.: Физматлит, 1963. — 76 с.
- 18) 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).
- 19) Kaufman L., Rousseeuw P.J. Finding Groups in Data: an introduction to cluster analysis / L. Kaufman, P.J. Rousseeuw. — Wiley, 1990 – 368 с.
- 20) 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.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ТЕРМИНОЛОГИЯ

Ниже приведен список необходимых терминов для ознакомления [15], [16], [17].

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.

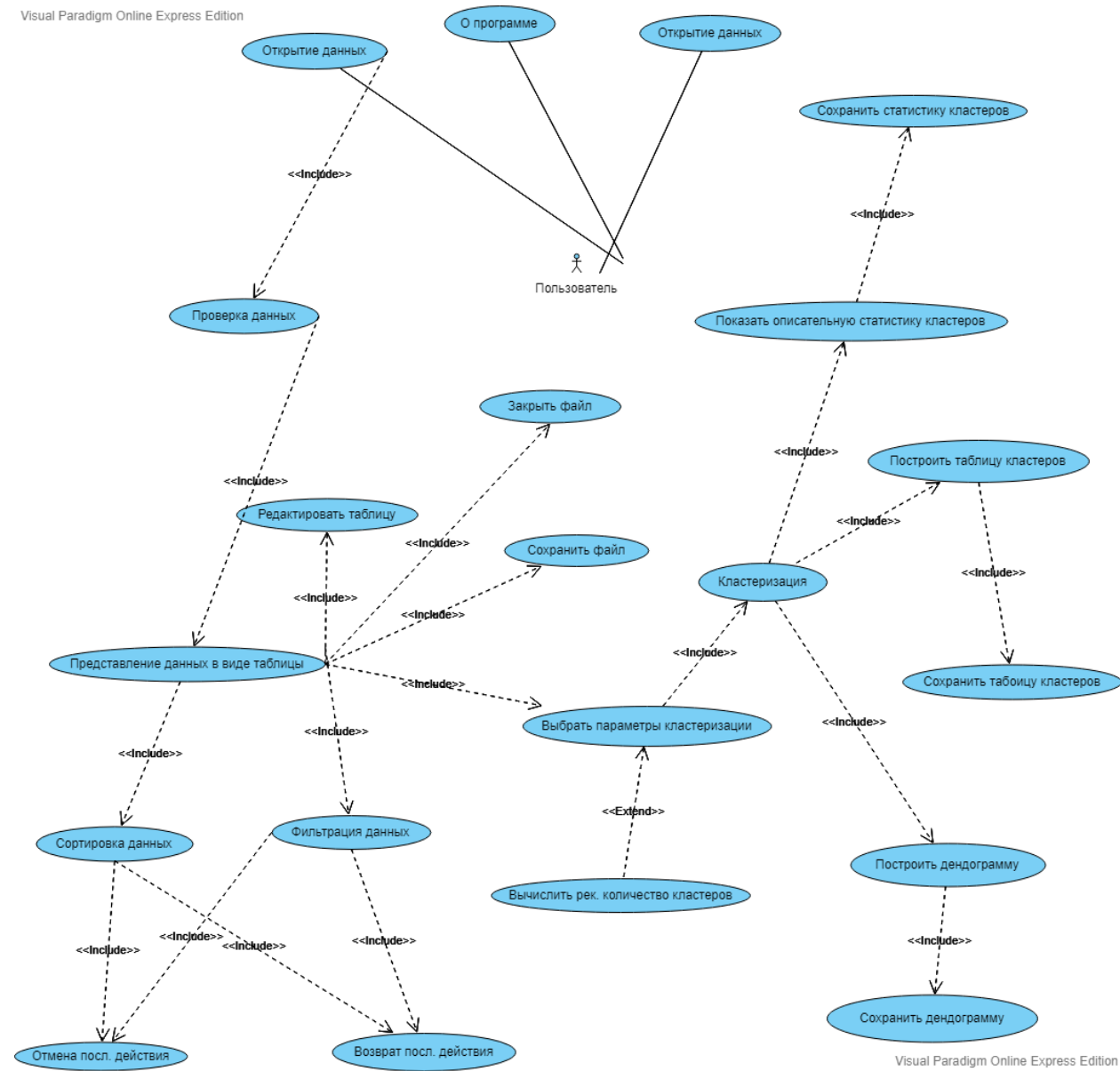
Стратегия объединения — алгоритм объединения двух кластеров.

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.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ДИАГРАММА ПРЕЦЕДЕНТОВ



Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 3

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.17701729.04.13-01 ТЗ				
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.17701729.04.13-01 ТЗ				
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.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ПРИЛОЖЕНИЕ 6

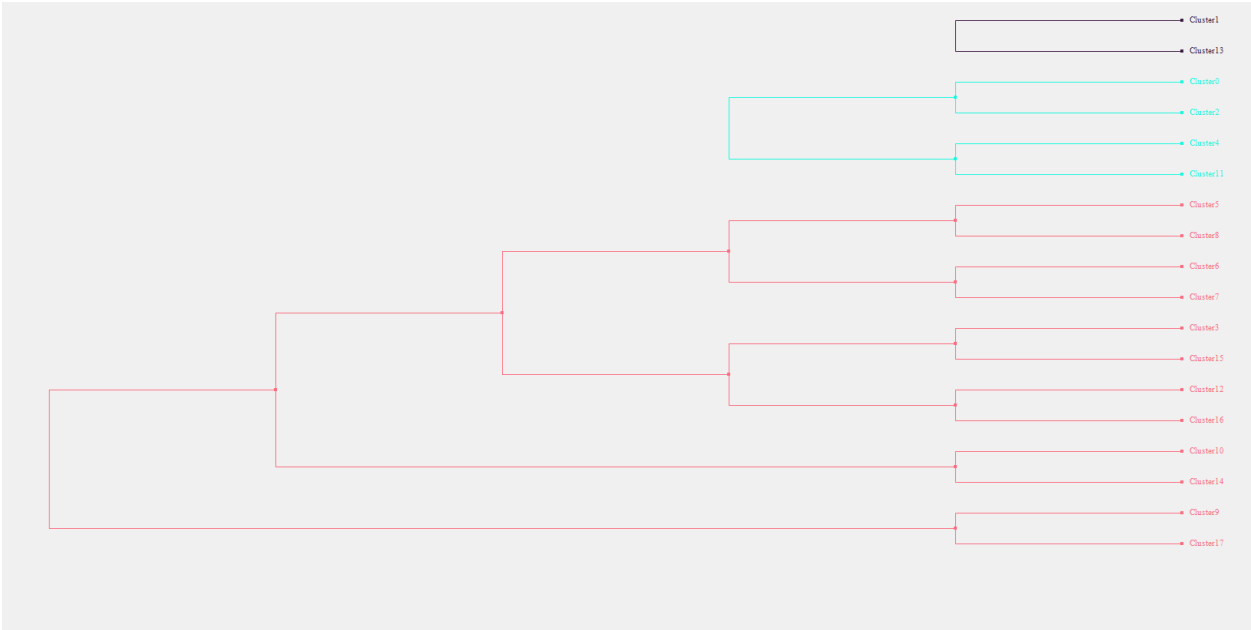
ПРИМЕР ВЫХОДНОГО ФАЙЛА ОПИСАТЕЛЬНОЙ СТАТИСТИКИ КЛАСТЕРОВ

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
3333333333;82.6308333333333;12

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
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.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

[illegible]

Изм.	Лист	№ докум.	Подп.	Дата
RU.17701729.04.13-01 ТЗ				
Инв. № подл.	Подп. и дата	Взам. инв. №	Инв. № дубл.	Подп. и дата