**Аннотация**

**Введение**

**1. Описание и анализ сферы разработки**

1.1 Описание и анализ предметной области

1.2 Описание алгоритмов анализа данных

**2. Разработка модулей системы и алгоритмов анализа данных**

2.1 Разработка алгоритмов анализа данных

2.1.1 Дерево принятия решений

2.1.2 Метод экспоненциального сглаживания

2.1.3 Средняя оценка

2.2 Разработка приложения

2.2.1 Описание архитектуры приложения

2.2.2 Библиотека доступа к хранилищу данных

2.2.3 Аналитическая библиотека

2.2.4 Веб проект

2.3 Разработка хранилища данных

**3. Практические возможности системы**

3.1 Просмотр информации о высшем учебном заведении

3.2 Выполнение классификации специальностей

3.3 Построение прогноза проходного балла

3.4 Общая оценка высших учебных заведений

**4. Соответствие научных положений исследования**

ВВЕДЕНИЕ

Интеллектуальный анализ данных – это обработка информации и выявление в ней моделей и тенденций, которые помогают принимать решения.

Алгоритм интеллектуального анализа данных — это набор эвристики и вычислений, который создает модель интеллектуального анализа данных из данных. Чтобы создать модель, алгоритм сначала анализирует предоставленные данные, осуществляя поиск определенных закономерностей и тенденций. Алгоритм использует результаты этого анализа для выбора оптимальных параметров создания модели интеллектуального анализа данных. Затем эти параметры применяются ко всему набору данных, чтобы выявить пригодные к использованию закономерности и получить подробную статистику.

В процесс построения модели интеллектуального анализа данных входят следующие задачи:

* формулировка задачи;
* сбор и подготовка исходных данных;
* построение модели;
* исследование модели;
* публикация и обновление модели.

Интеллектуальный анализ данных применяется в различных сферах деятельности, в том числе и сфере образования. За последние годы высшими учебными заведениями собрано большое количество информации о результатах наборов студентов. Подвергнув данные анализу можно получить новые знания о специальностях для поступающих и сотрудников учебных заведений.

Целью работы является разработка веб–приложения и алгоритмов анализа специальностей высших учебных заведений.

Для достижения данной цели поставлены следующие задачи:

* осуществить сбор данных по статистике набора студентов высших учебных заведений;
* разработать хранилище данных и веб–приложение;
* выбрать и реализовать алгоритмы анализа данных;
* оценить результаты анализа на полноту и объективность.

В качестве методов исследования применяются методы интеллектуального анализа данных, а именно классификация и прогнозирование. Данные методы обеспечивают решение поставленной задачи.

Элементом научной новизны являетсяполучение новых знаний о специальностях высших учебных заведений по данным наборов студентов, а также централизованное хранение информации и результатов анализа.

Практической составляющей является разработанное веб-приложение, доступное в сети интернет, которое предоставляет общую информацию по каждому высшему учебному заведению и результаты интеллектуального анализа данных в визуальном представлении.

Разработка программного комплекса осуществляется в интегрированной среде разработке Visual Studio 2017 на платформе .NET Framework, с применением технологии ASP.NET Core 2 и языков программирования C#, JavaScript. В качестве хранилища данных применятся система управления базами данных MS SQL Server 2016.

Положения на защиту:

* алгоритм дерево принятия решений для классификации специальностей;
* алгоритм прогноза проходного балла на следующий год методом экспоненциального сглаживания;
* представление возможностей разработанного веб-приложения.

Результаты исследования опубликованы в следующих сборниках:

* Новые информационные технологии в научных исследованиях: материалы ХХ Юбилейной Всероссийской научно-технической конференции студентов, молодых ученых и специалистов// Рязанский государственный радиотехнический университет. 2015. 300с. Тезисы на тему «Разработка информационной системы анализа потребления жилищно-коммунальных услуг на основе хранилища данных» [2].

1. **Описание и анализ области разработки**

Развитие информационных технологий существенно повлияло на рост объемов данных. Большая часть информации не приносит пользы, поскольку человек не в состоянии переработать такое количество данных. В результате возникает проблема извлечения полезной информации для пользователя из большого объема данных. Для решения проблемы применяются методы интеллектуального анализа данных. Данная область является наиболее развивающейся в сфере информационных технологий и предназначена для определения полезных знаний из различных наборов данных.

В выбранной области исследования в процессе поиска существующих решений не было найдено разработок и решений, которые были направлены исключительно на получение новых знаний о специальностях высших учебных заведений с применением методов интеллектуального анализа данных. Были изучены различные информационные ресурсы с общей статистикой и описанием специальностей, но без применения аналитики, что не позволяет получить полную оценку.

Данная работа посвящена разработке системы интеллектуального анализа данных сфере образования, а именно анализ специальностей высших учебных заведений по данным итогов набора студентов. Система включает в себя хранилище данных, алгоритмы методов интеллектуального анализа данных, пользовательский интерфейс и визуализацию результатов анализа.

* 1. **Анализ области разработки**

Интеллектуальный анализ данных - мультидисциплинарная область, возникшая и развивающаяся на базе таких наук, как прикладная статистика, распознавание образов, искусственный интеллект, теория баз данных. Интеллектуальный анализ данных представляет собой процесс поддержки принятия решений, основанный на поиске в данных неочевидных, объективных и полезных на практике закономерностей [].

Основными задачами интеллектуального анализа данных являются:

* прогнозирование;
* классификация;
* ассоциация;
* визуализация;
* кластеризация.

Прогнозирование осуществляется на основе исторических данных, а в результате оцениваются будущие или пропущенные численные значения. Данная задача реализуется с применением методов математической статистики, нейронных сетей.

Классификация определяет свойства, которые характеризуют общие группы объектов из исходного набора данных. По данному анализу можно выяснить к какой группе относится объект.

Кластеризация является более сложной задачей классификации, поскольку классы объектов изначально не предопределены. Результатом данной задачи интеллектуального анализа данных является разбиение объектов на группы.

Задача ассоциации выполняет поиск закономерностей между связанными событиями в исходном наборе данных. Ассоциация является уникальной задачей интеллектуального анализа данных так, как поиск закономерностей осуществляется между несколькими событиями, которые происходят одновременно, а не на основе свойств анализируемого объекта.

Визуализация решает задачу графического представления анализируемых данных, используя графические методы, показывающие наличие закономерностей в данных.

Процесс реализации интеллектуальный анализ данных состоит из следующих этапов:

1. Постановка задачи анализа данных и анализ предметной области.
2. Сбор сведений и данных.
3. Обработка данных может включать в себя следующие операции: очистку, интеграцию, преобразование данных.
4. Применение алгоритмов интеллектуального анализа.
5. Описание результатов анализа в графическом, числовом или другом виде.
6. Практическое применение новых знаний.

Для практического применения разрабатываются системы интеллектуального анализа данных, которые состоят из следующих компонентов:

1. Хранилище данных. В качестве хранилища данных могут применятся реляционные базы данных, объектно-реляционные базы данных и другие хранилища.
2. Сервер хранилища данных осуществляет извлечение данных на основании запросов из пользовательского интерфейса.
3. База знаний о предметной области, которая указывает, как проводить поиск и оценивать полезность результатов анализа.
4. Методы интеллектуального анализа по поиску новых знаний.
5. Графический пользовательский интерфейс. Данный модуль обеспечивает взаимодействие пользователя и системы интеллектуального анализа данных. Пользователь получает результаты анализа в графическом виде.

Основная задача системы - это получение новых знаний о специальностях высших учебных заведений с применением методов интеллектуального анализа. Исходными данными являются открытые сведения о итогах набора студентов за прошедшие годы, размещенные на веб-сайтах университетов. Для хранения информации о предметной области, результатов анализа используется реляционная база данных. Методы интеллектуального анализа реализованы в отдельном программном модуле и входят в состав веб-приложения, которое выполняет визуализацию результатов и обеспечивает взаимодействие системы с пользователем.

В ходе анализа предметной области и исходных данных были выбраны основные методы анализа, а именно метод классификации и прогнозирования. С их помощью достигается классификация специальностей по группам и прогноз проходного балла на будущий год. Выше перечисленные методы в полном объеме решают поставленные задачи по получению новых знаний.

Классификация – это процесс упорядочивания по некоторому принципу множества объектов, которые имеют признаки для определения сходства или различия между этими объектами. Нахождение моделей или функций, которые описывают и различают классы, позволяет предсказывать класс произвольного заданного объекта с известными атрибутами, но неизвестной меткой класса.

Классификация может быть одномерной и многомерной.

Основным требованием классификации является наличия признаков, которые характеризуют группу принадлежности объекта. Задачей этого метода обычно называют определение категории зависимой переменной по выборке категориальных переменных. Процесс классификации состоит из следующих этапов: конструирования модели и ее использования. Использование заключается в классификации новых знаний. Известные знания сравниваются с результатами использования новой модели.

Например, в настоящей работе определяются группы, в которые попадают специальности учебных заведений на основе информации о проходных баллах, количества заявлений, количества выделенных мест.

Точность оценки классификации достигается за счет процедуры оценки точности классификации на данных из тестового множества. Новая модель считается корректной при условии, если тестовое множество имеет примерно такие же результаты, как и обучающее. Получение тестовых и обучающих выборок происходит за счет пропорционального деления исходных данных.

Существует перечень различных методов и алгоритмов для реализации классификации, например, следующие:

* методом опорных векторов;
* деревьев принятия решений;
* линейная регрессия;
* байесовская классификация.

Выбор метода классификации для применения обусловлен следующими свойствами:

* скорость создания модели;
* устойчивость к нарушениям;
* надежность;
* интерпретируемость.

Прогнозированием является определение функциональной зависимости между зависимыми и независимыми значениями. Главной целью данного метода предсказать будущие значения или явления.

Для реализации прогноза необходима обучающая выборка данных. В настоящей работе прогноз осуществляется по информации проходных баллов специальности, которая хранится в виде временного ряда. Временным рядом называется последовательность наблюдаемых значений, упорядоченные в определенные моменты времени.

Главными свойствами временного ряда выступают:

* тренд;
* сезонная составляющая;
* цикличность.

Трендом называют неслучайную функцию, которая формируется под действием общих или долговременных тенденций, влияющих на временной ряд. Сезонная компонента временного ряда является периодически повторяющейся составляющей временного ряда. **Цикличностью**называются периодические колебания, выходящие за рамки одного года [].

Определение ошибки прогноза ключевым образом зависит от метода реализации. Существуют следующие виды ошибок:

1. Среднеквадратическая ошибка является суммой квадратов ошибок.
2. Средняя ошибка является средним значением ошибок.
3. Относительная ошибка.
4. Средняя абсолютная ошибка.

Периодом прогноза есть единица времени, на которую выполняется прогноз. В данной работе используется краткосрочный прогноз, выбор обусловлен редкому пополнению исходных данных и спецификой предметной области. Прогноз осуществляется на один шаг вперед, что позволяет применить статистический метод анализа. Для получения достоверного и корректного прогноза исходные данные должны быть точными и достоверными.

* 1. **Описание алгоритмов анализа данных**

# В данной области исследования применяется два метода интеллектуального анализа данных: дерево принятия решений для классификации специальностей и прогнозирование проходного балла на основе метода экспоненциального сглаживания.

# Метод деревьев решений является одним из наиболее популярных методов решения задач классификации. Деревья решений позволяют определить принадлежность объектов к тому или иному классу в зависимости от принимаемого значения. Например, специальность имеющая проходной балл 180 баллов будет в одной группе, а специальность с 210 баллами в другой.

# В общем представление дерево является последовательной структурой, которая представляет набор иерархических правил. Конечные узлы дерева считаются решением классификации, а внутренние узлы дерева выполняют проверку заданного условия. В результате прохождения от корня дерева до его вершины решается задача классификации, выбирается один из классов. Целью построения дерева решения в нашем случае является определение значения категориальной зависимой переменной.

# 

# Рисунок 1 – Структурная схема дерева решений

# При проектировании модели анализа данных создается набор правил, по которым осуществляется построение дерева. Правилами узлов дерева является логическая операция ветвления.

# Также стоит выделить, что дерево состоит из атрибутов и классов. Атрибуты – это данные предметной области процесса (проходной балл, количество заявлений), а класс итоговая группа классификации. Совокупность информации о атрибутах в одном конкретном узле называется критерием расщепления. Корректность и качество дерева решений зависит от выбранного критерия расщепления.

# В ходе создания дерева решаются вопросы выбора критерия расщепления и остановки обучения. В ходе этапа сокращения дерева решается вопрос отсечения некоторых его ветвей. Процесс создания дерева происходит сверху вниз, т.е. является нисходящим. В ходе процесса алгоритм должен найти такой критерий расщепления, иногда также называемый критерием разбиения, чтобы разбить множество на подмножества, которые бы ассоциировались с данным узлом проверки. Каждый узел проверки должен быть помечен определенным атрибутом. Существует правило выбора атрибута: он должен разбивать исходное множество данных таким образом, чтобы объекты подмножеств, получаемых в результате этого разбиения, являлись представителями одного класса или же были максимально приближены к такому разбиению. Существуют различные критерии расщепления. В некоторых методах для выбора атрибута расщепления используется так называемая мера информативности подпространств атрибутов, которая основывается на энтропийном подходе и известна под названием мера энтропии. Другой критерий расщепления реализован в алгоритме CART и называется индексом Gini. При помощи этого индекса атрибут выбирается на основании расстояний между распределениями классов.

# В процессе построения дерева, чтобы его размеры не стали чрезмерно большими, используют специальные процедуры, которые позволяют создавать оптимальные деревья, так называемые деревья подходящих размеров.

# Выбор реализации алгоритма классификации методом дерева решений обусловлен следующими преимуществами:

# Позволяет работать с дискретными и непрерывными значениями, что обеспечивает большую гибкость.

# Для реализации дерева решений не требуется больших объемов данных.

# Наглядность.

# Высокая скорость обучения.

# Данный алгоритм не требует самостоятельного выбора основных атрибутов анализа, он выбирает их самостоятельно.

# Однако данный алгоритм имеет недостатки, одним из которых является проблема усечения узлов у больших деревьев.

# В данном исследовании для решения задачи классификации выбран алгоритм дерева принятия решений С4.5. Выбранный алгоритм решает исключительно задачи классификации и не имеет ограничений по количеству потомков у узлов. Далее, рассматривается математическая реализация данного подхода.

# РАСЧЕТ И БЛОК СХЕМА

Для прогнозирования проходного балла на следующий год выбран метод экспоненциального сглаживания, который применятся для сглаживания временных рядов. Выбор данного метода обусловлен тем, что выполняется краткосрочный прогноз, что обеспечивает получить достоверный и корректный прогноз, а также простота реализации.

Главной идеей рассматриваемого метода является вычисление экспоненциальных скользящих средних для ряда, содержащего проходные баллы определенной специальности за все год, которые находятся в базе данных. Преимуществом экспоненциального сглаживания заключается в простой адаптации при возникновении новых значений, что обеспечивает надежность прогноза.

Расчет данного метода осуществляется с применением следующих формул:

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsm_growth_t2_1.gif

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsmooth_1.gif, где

*Xt*-1 - фактическое наблюдение в момент *t*-1;

*St* - значение экспоненциального среднего в момент *t*;

*α* - параметр сглаживания, *α*= *const*, *α*ϵ (0; 1].

Экспоненциальное среднее в момент *t* здесь выражено как взвешенная сумма текущего наблюдения и экспоненциального среднего прошлого наблюдения с весами *α* и (1 - *α*) соответственно. При рекуррентном использовании данного подхода значение *St* выражается через *X* и формула приводится к виду:

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsmooth_2.gif

Следовательно, величина *St* является взвешенной суммой членов ряда и значение веса уменьшается экспоненциально в зависимости от удаленности наблюдения относительно момента *t*. Метод экспоненциального сглаживания на вход принимает исходный ряд, а на выходе формирует значения экспоненциальных средних. В исследуемой работе выходным значением является результат прогноза на будущий год.

Сглаженный ряд *St* имеет тоже математическое ожидание, что и ряд *X*, но меньшую дисперсию. При высоком значении *α* дисперсия сглаженного ряда не значительно отличается от дисперсии ряда *X*. Чем меньше *α*, тем в большей степени сокращается дисперсия сглаженного ряда (то есть подавляются колебания исходного ряда).

Далее экспоненциальное среднее можно использовать для построения краткосрочных прогнозов. В этом случае предполагается, что исходный ряд описывается моделью:

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsmooth_3.gif, где

*at* - изменяющийся во времени средний уровень ряда;

*errt* - случайные отклонения с нулевым математическим ожиданием.

Прогнозная модель имеет вид:

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsmooth_4.gif, где

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsmooth_5.gif - прогноз, сделанный в момент *T* на τ единиц времени вперед;

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsmooth_6.gif - оценка.

Оценкой параметра модели *aT* служит экспоненциальное среднее ряда *ST*. Таким образом, все свойства экспоненциального среднего распространяются на прогнозную модель. В частности, если привести рекуррентную формулу к следующему виду:

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsmooth_7.gif

и рассматривать *St*-1 как прогноз на один шаг вперед, то величина (*Xt*-1 - *St*-1) есть погрешность этого прогноза, а новый прогноз *St* получается в результате корректировки предыдущего прогноза с учетом его ошибки. В этом и состоит сущность адаптации.

На основе простого экспоненциального сглаживания были разработаны более сложные модели сглаживания временных рядов, содержащих периодические сезонные колебания и/или обладающих тенденцией роста.

Данная система позволяет строить наряду с простым экспоненциальным сглаживанием модели, отражающие эффекты роста (линейного, экспоненциального или затухающего) и сезонности (аддитивного или мультипликативного), которыми обладает исходный ряд.

В общем виде рекуррентная формула экспоненциального сглаживания записывается следующим образом:

http://help.prognoz.com/ru/mergedProjects/Lib/img/expsmooth_9.gif, где

множители *d*1 и *d*2 определяются в зависимости от выбранной модели сглаживания.

**2. Разработка модулей системы и алгоритмов анализа данных**

**2.1 Разработка алгоритмов анализа данных**

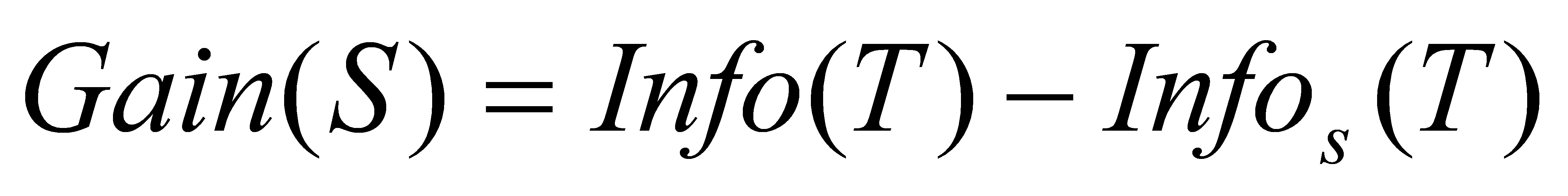
**2.1.1 Дерево принятия решений**

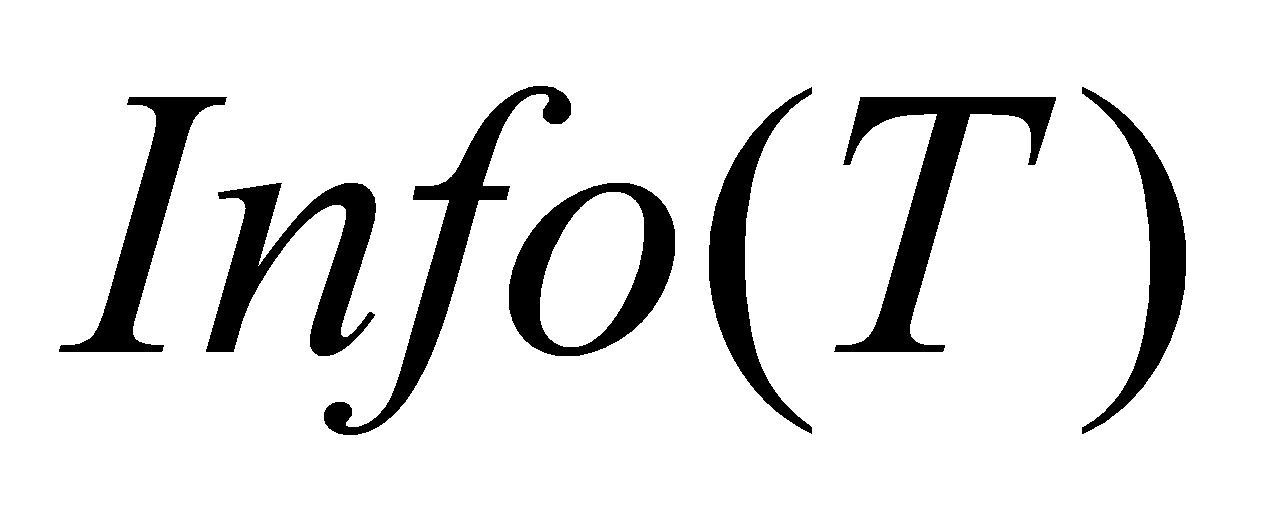
Работы алгоритма рассматривается по данным Рязанского государственного радиотехнического университета за 2016 год.

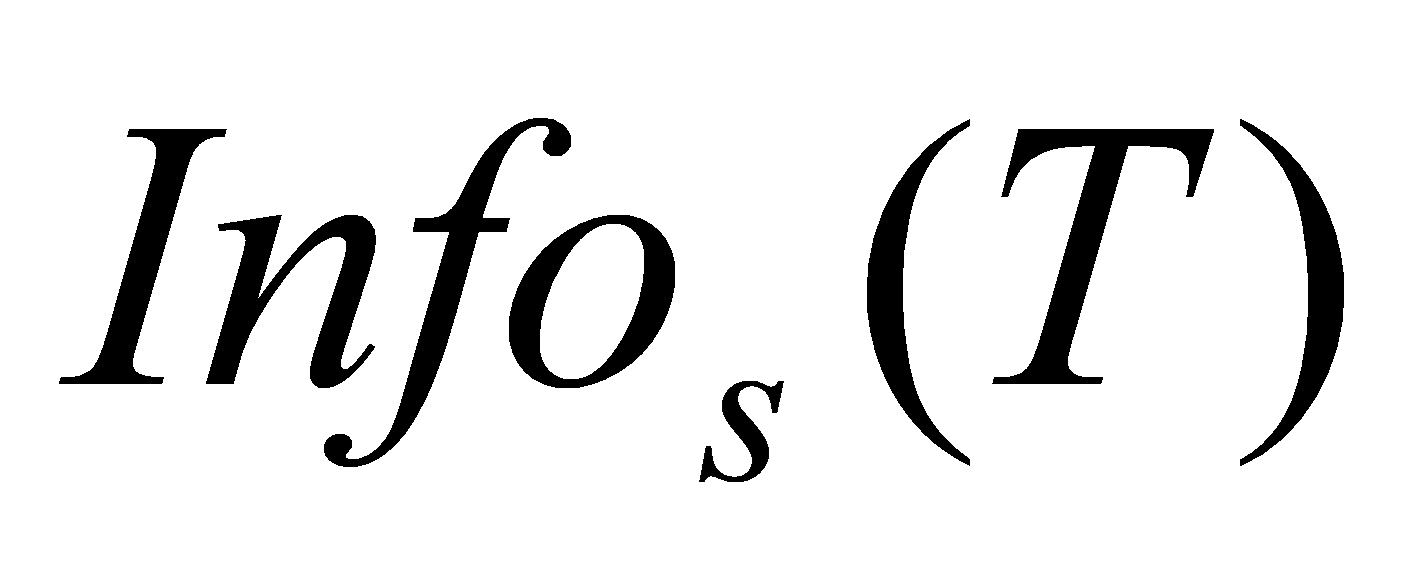
Данные факультета автоматики и информационных технологий в управлении

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Направление обучения* | *Проходной балл* | *Количество мест* | *Количество заявлений на поступление* | *Конкурс* | *Зачислено на платной основе* |
| Управление в технических системах | 186,96 | 32 | 175 | 5,47 | 3 |
| Информационные системы и технологии | 198,8 | 20 | 113 | 5,65 | 3 |
| Стандартизация и метрология | 202,5 | 10 | 38 | 3,8 | 0 |
| Автоматизация технологических процессов и производств | 202,56 | 32 | 186 | 5,81 | 7 |
| Биотехнические системы и технологии | 181 | 10 | 70 | 5,81 | 0 |
| Приборостроение | 188,75 | 15 | 65 | 4,33 | 0 |
| Мехатроника и робототехника | 187,71 | 14 | 102 | 7,29 | 1 |

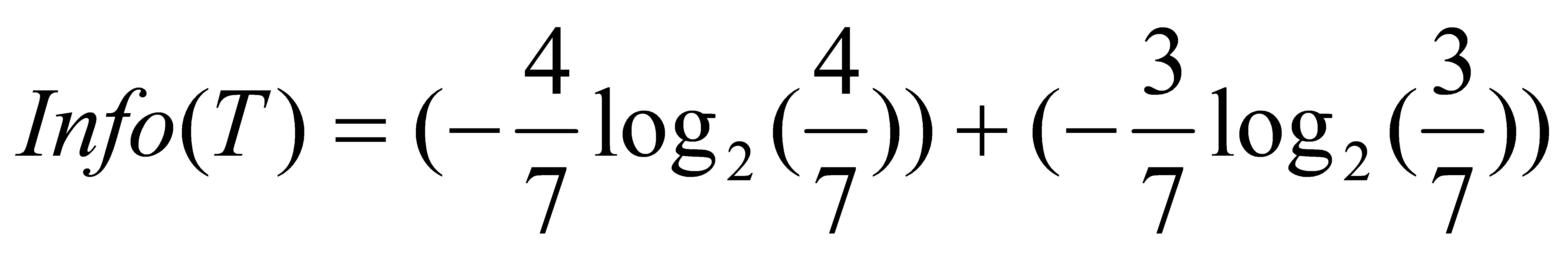
В данном случае для реализации алгоритма атрибутами являются проходной балл, количество мест, количество заявлений на поступление. На следующем шаге рассчитывается прирост информации каждого атрибута, по формуле:

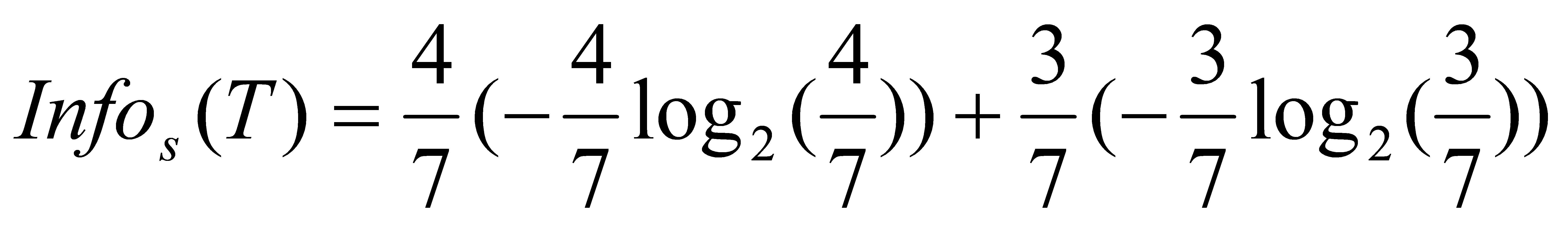
, где

— энтропия множества T до разбиения;

 — энтропия после разбиения S.

Вычисление прироста информации атрибута проходной балл:





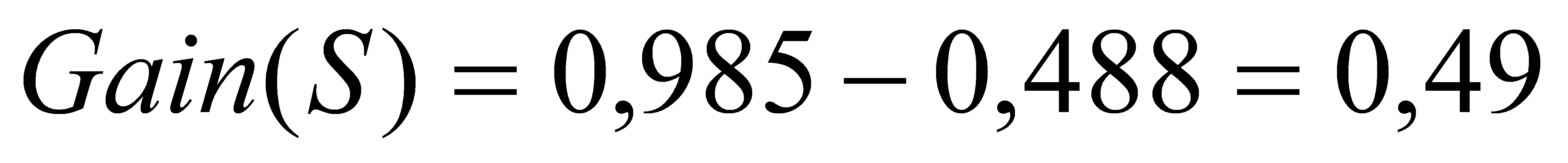
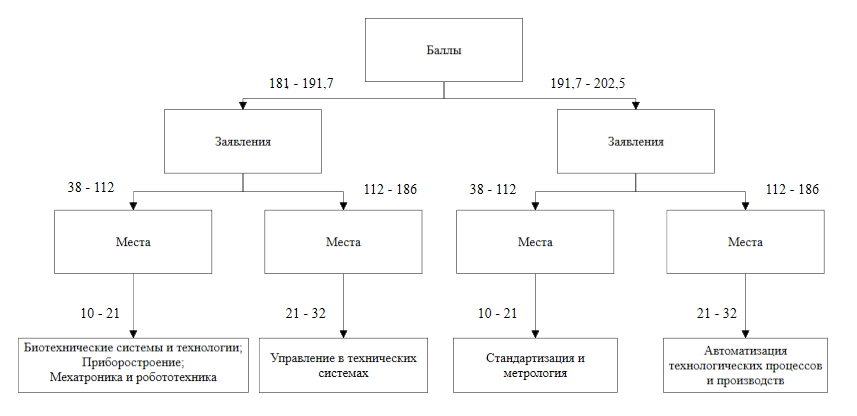


Таблица 2

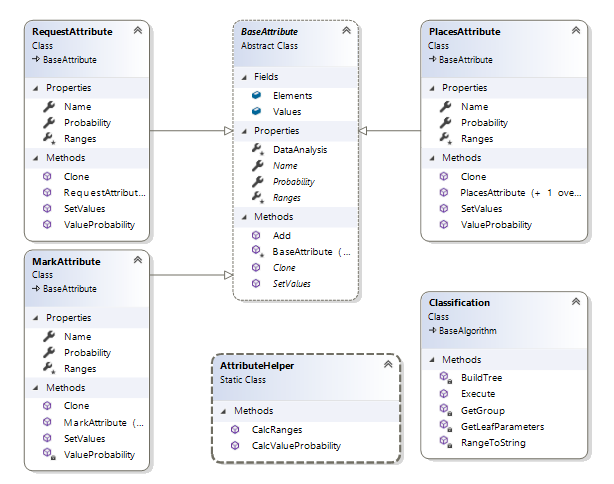
**Описание атрибутов**

|  |  |  |
| --- | --- | --- |
| *Атрибут* | *Прирост информации* | *Условие классификации* |
| Проходной балл | 0,49 | (181 – 191,7) и (191,7 – 202,5) |
| Количество мест | 0.49 | (10 – 21) и (21 – 32) |
| Количество заявлений на поступление | 0.46 | (38 – 112) и (112 – 186) |

Результат классификации направлений факультета автоматики и информационных технологий в управлении представлен на рисунке 3.



**Рис. 3 – Дерево решений**



**2.1.2 Метод экспоненциального сглаживания**

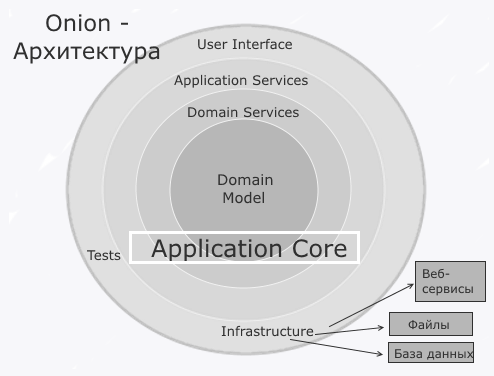
**2.1.3 Средняя оценка**

**2.2 Разработка приложения**

Разработка программного обеспечения по данному исследованию велась в интегрированной среде разработке Visual Studio 2017 с применением возможностей программной платформы .Net Framework 4.6. Для создания веб-приложения используется технология ASP.NET Core 2. Основным языком программирования для разработки серверной части приложения и библиотек проекта является С#, а для создания клиентской части TypeScript и Angular 4.

**2.2.1 Описание архитектуры приложения**

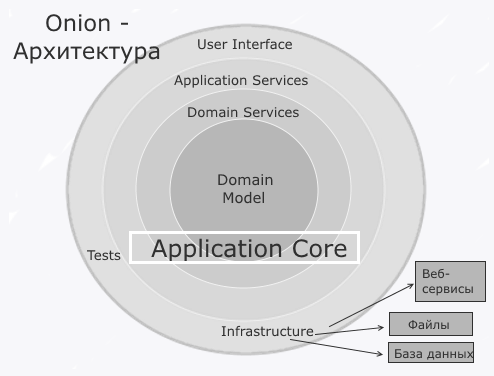
Данный проект имеет многоуровневую архитектуру состоит из трех основных модулей: веб-проект, модуль доступа к базе данных, модуль алгоритмов интеллектуального анализа данных.



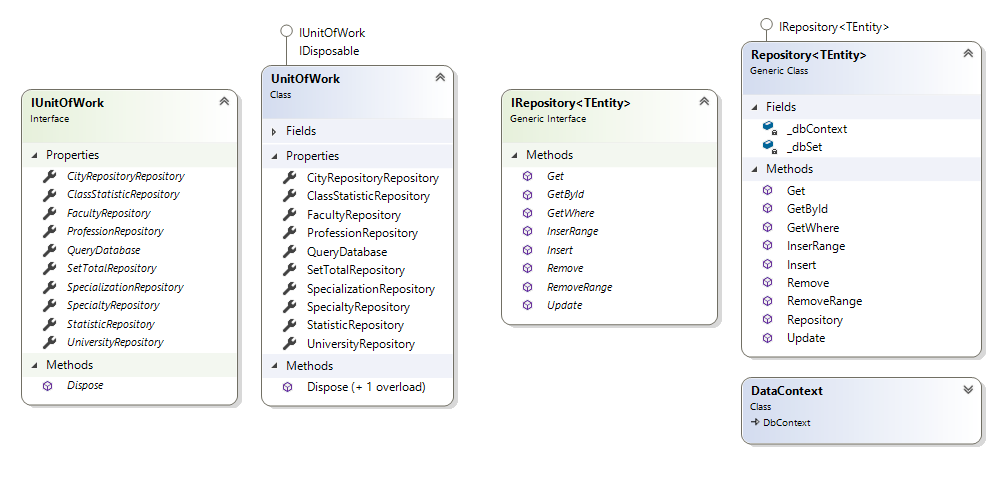
Это архитектурное решение имеет ряд преимуществ:

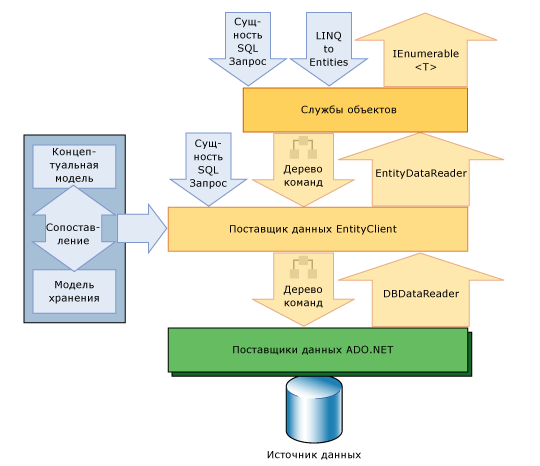
1. Обеспечивается гибкость так, как каждый уровень может использоваться и функционировать отдельно, что позволяет вносить изменения в отдельные модули, а не в систему целом.
2. Масштабируемость, которая достигается за счет развертывания слоев системы.
3. В системе можно применять модульную архитектуру, которая позволяет использовать отдельные компоненты, что повышает доступность.

Для реализации веб-проекта используется шаблон архитектуры MVC (Model-View-Controller). Эта архитектура разделяет веб решение на три части: модель, представление, контроллер. Представление является реализацией пользовательского интерфейса и реагирует на все действия пользователя. Модель реализует бизнес-логику приложения. Контроллер выполняет роль связующего звена между моделью и представлением. Например, с представления поступает запрос в контроллер на получение данных модели, в свою очередь, контроллер принимает запрос и обращается к модели, а затем возвращает ответ на представление.

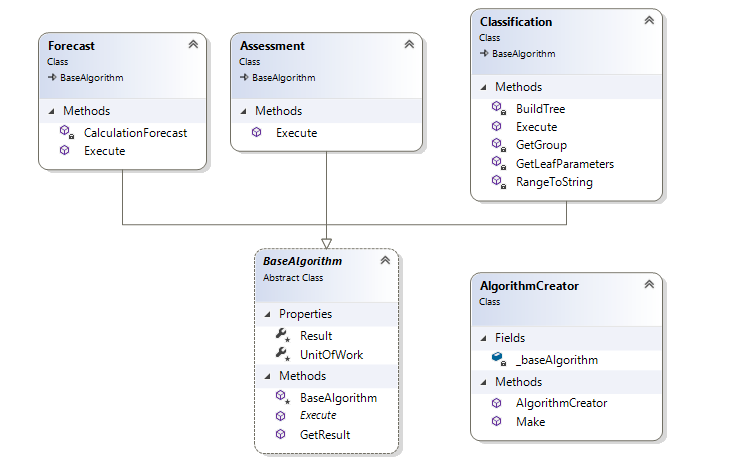


**2.2.2 Библиотека доступа к хранилищу данных**





**2.2.3 Аналитическая библиотека**



public class AlgorithmCreator

{

private BaseAlgorithm \_baseAlgorithm;

public AlgorithmCreator(BaseAlgorithm baseAlgorithm)

{

\_baseAlgorithm = baseAlgorithm;

}

public void Make(IDictionary<string, string> parameters = null)

{

\_baseAlgorithm.Execute(parameters);

}

}

public abstract class BaseAlgorithm

{

protected string Result { get; set; }

protected IUnitOfWork UnitOfWork { get; private set; }

protected BaseAlgorithm()

{

UnitOfWork = new UnitOfWork();

}

public virtual string GetResult()

{

return Result;

}

public abstract void Execute(IDictionary<string, string> parameters = null);

}

**2.3 Разработка хранилища данных**

Для централизованного хранения данных о высших учебных заведениях и результатах аналитики разработано единое хранилище. Разработка хранилища данных осуществляется с помощью системы управления реляционными базами данных MS SQL. Данная система применятся для работы с персональными и крупными базами данных. Основным языком программирования является Transact–SQL.

База данных состоит из девяти таблиц и содержит информацию о высших учебных заведениях, факультетах, специальностях, городах, специализациях, итогам наборов и результатах анализа. Далее представлена детальная информация о каждой таблице базы данных и ее структурная схема рисунок 1.

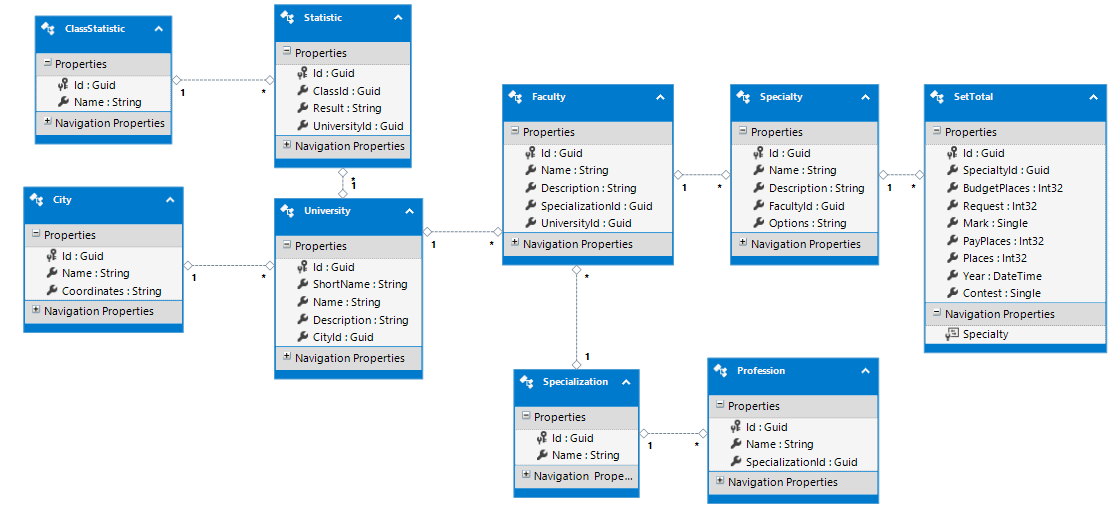


Рисунок – Структурная схема базы данных

Таблица ClassStatistic является справочником методов интеллектуального анализа. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы ClassStatistic

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ, идентификатор |
| Name | Nvarchar(100) | Наименование класса интеллектуального анализа данных |

Таблица Statistic содержит данные о результатах анализа по каждому университету. Данная таблица имеет связи один ко многим с таблицами University и ClassStatistic. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы «Statistic»

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ, идентификатор |
| ClassId | Guid | Внешний ключ, идентификатор класса анализаданных |
| UniversityId | Guid | Внешний ключ, идентификатор университета |
| Result | Json | Результаты анализа данных |

Таблица City является справочником городов. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы «City»

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ |
| Name | Nvarchar(50) | Наименование города |
| Coordinates | Json | Координаты |

Таблица University содержит данные о высших учебных заведениях. Данная таблица имеет связь один ко многим с таблицей City. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы «University»

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ, идентификатор |
| ShortName | Nvarchar(30) | Аббревиатура |
| Name | Nvarchar(100) | Наименование |
| Description | Nvarchar(max) | Описание |
| CityId | Guid | Внешний ключ, идентификатор города |

Таблица Faculty содержит данные о факультетах университетов. Данная таблица имеет связи один ко многим с таблицами Specialization и University. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы «Faculty»

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ, идентификатор |
| Name | Nvarchar(100) | Наименование |
| Description | Nvarchar(max) | Описание |
| SpecializationId | Guid | Внешний ключ, идентификатор специализации |
| UniversityId | Guid | Внешний ключ, университета |

Таблица Specialization является справочником специализаций. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы «Specialization»

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ, идентификатор |
| Name | Nvarchar(100) | Наименование специализации |

Таблица Profession является справочником профессий. Данная таблица имеет связь один ко многим с таблицей Specialization. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы «Profession»

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ, идентификатор |
| Name | Nvarchar(100) | Наименование специализации |
| SpecializationId | Guid | Внешний ключ, идентификатор специализации |

Таблица Specialty содержит данные о специальностях по каждому факультету. Данная таблица имеет связь один ко многим с таблицей Faculty. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы «Specialty»

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ, идентификатор |
| Name | Nvarchar(100) | Наименование |
| Description | Nvarchar(max) | Описание |
| Options | Json | Опции |
| FacultyId | Guid | Внешний ключ, идентификатор факультета |

Таблица SetTotal содержит данные по результатам наборов за каждый год. Данная таблица имеет связь один ко многим с таблицей Specialty. Описание таблицы представлено в таблице .

Таблица 10 – Описание таблицы «SetTotal»

|  |  |  |
| --- | --- | --- |
| Наименование поля | Тип данных | Описание |
| Id | Guid | Первичный ключ, идентификатор |
| SpecialtyId | Guid | Внешний ключ, идентификатор специальности |
| BudgetPlaces | Integer | Количество бюджетных мест |
| Request | Integer | Количество поданных заявлений |
| Mark | Real | Проходной балл |
| PayPlaces | Integer | Количество платных мест |
| Places | Integer | Количество мест |
| Year | DateTime | Год |
| Contest | Real | Конкурс на специальность |

**Список литературы**

1. Глебова Л.Н., Кузнецова М.Д. Мониторинг качества высшего педагогического образования. М.: Логос, 2012. 368 с.

2. Чубарова О.И. Образовательный риск как экономическая категория, его сущность // Пол- зуновский вестник. 2005. № 1. С. 199–208. 3. XIN Tao, LI Feng & LI Ling-yan An International Comparison of Elementary Education Quality Assessment // Journal of Beijing Normal University (Social Science Edition). 2007. P. 5–10.

4. Brown G. The Quality of Learning of Learning and How to Assess it // Lifelong Learning in Europe. 1999. Vol. 4. Issue 1. P. 47–54.

5. Dohmen G. Lifelong Learning for All – innovative perspectives of continuing education // Lifelong Learning in Europe. 1999. Vol. IV. Issue 3. P. 154–158.

6. Dinham S. The quality teaching movement in Australia encounters difficult terrain: a personal perspective // Australian Journal of Education. 2013. № 57. P. 91–106. DOI: 10.1177/0004944113485840.

7. Гуров В.С., Корячко В.П., Таганов А.И., Моисеенко В.П., Таганов Р.А. Опыт создания и применения ресурсов электронной информационно-образовательной среды по направлению ИПИ (CALS) и CASE (САПР) – технологий // Труды VII Всероссийской научно-методической конференции «Телематика-2010». 2010. Том 1. С. 166–167.

8. Heyneman S.P., Lee B. The impact of international studies of academic achievement on policy and research // Handbook of international large-scale assessment: Background, technical issues and methods of data analysis. 2014. P. 37–72.

9. Kamens D.H., McNeely C.L. Globalization and the growth of international educational testing and national assessment. Comparative Education Review. 2009. № 54. P. 5–25. DOI: 10.1086/648471.

10. Таганов А.И., Таганов Р.А. Методические основы создания информационных систем сфе- ры образования. Рязань: РГРТУ, 2006. 342 с.

11. Attfield I., Vu B.T. A rising tide of primary school standards: The role of data systems in improving equitable access for all to quality education in Vietnam // International Journal of Educational Development. 2013. № 33. P. 74–87. DOI: 10.1016/j.ijedudev.2012.02.003.

12. Колесенков А.Н. Таганов А.И. Концепция геоинформационной технологии мониторинга образовательных программ онлайн-обучения // Открытое и дистанционное образование. 2015. Вып. № 60. С. 69–73.

13. Колесенков А.Н. Информационная под- держка принятия решений при организации и построении систем дистанционного обучения // Дистанционное и виртуальное обучение. 2016. № 5 (107). С. 62–68.

14. Колесенков А.Н., Таганов А.И. Примене- ние геоинформационных технологий в задачах управления образовательным процессом вуза // Управление качеством инженерного образова- ния. Возможности ВУЗов и потребности про- мышленности: Тезисы докладов второй между- народной научно-практической конференции. 2016. С. 121.

15. Костров Б.В., Баранчиков А.И. Теория и методы исследования моделей и алгоритмов представления данных для предметных облас- тей с ранжируемыми атрибутами // Вестник РГ- РТУ. 2013. № 5. выпуск 47. С. 59–64.

16. Барсегян А., Куприянов М., Холод И., Тесс М., Елизаров С. Анализ данных и процес- сов. СПб: БХВ-Петербург, 2009, 544 с.

17. Балабанов И.Т. Риск-менеджмент. М.: ФиС, 1996, 192 с.

18. Чернова Г.В., Кудрявцева А.А. Управление рисками. М.: Из-во Проспект, 2003. 160 с.

19. Вяткин В.Н., Вяткин И.В., Галуза В.А. Риск менеджмент. М.: Дашков и К, 2003, 512 с.

20. Саркисян С.А. Теория прогнозирования и принятия решений. М: Высшая школа, 1977, 351 с.