

ФИЗИКО-МАТЕМАТИЧЕСКИЕ НАУКИ

УДК 519.688

ИНФОРМАЦИОННАЯ СИСТЕМА МНОГОФАКТОРНОГО АНАЛИЗА УЧЕБНОЙ ДЕЯТЕЛЬНОСТИ ОБРАЗОВАТЕЛЬНОЙ ОРГАНИЗАЦИИ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ BIGDATA

© П.В. Агапов

Аннотация. Приведены основные результаты разработки структуры данных и программного средства для анализа образовательного процесса в Тамбовском государственном университете им. Г.Р. Державина. В качестве исходных данных используются обезличенные сгенерированные табличные значения. Описаны основные цели и задачи проекта, его функциональность, а также реализация интеллектуальной платформы с использованием языка объектно-ориентированного программирования Python. В дальнейшем планируется расширение функциональности разработанного программного средства.

Ключевые слова: анализ данных, базы данных, большие данные, программные средства, Python, MySQL

ВВЕДЕНИЕ

В настоящее время наблюдается стремительный рост объема информации во всех сферах жизни человека. Анализ разнотипных и неструктурированных данных является ключевой задачей цифровизации современного общества. Технологии хранения и масштабирования больших данных, алгоритмы их обработки составляют основу нового направления науки BigData.

Большие данные (BigData) – это комбинация структурированных, полуструктурированных и неструктурированных данных, собранных для того, чтобы использоваться в проектах машинного обучения, прогнозного моделирования, и аналитики собранных данных для получения информации о различных критериях эффективности ведения деятельности различных организаций [1].

В последнее время технологии BigData стали применяться для анализа факторов, влияющих на образовательный процесс. Аналитика данных и накопленная статистическая информация становятся неотъемле-

мой частью принятия управленческих решений в сфере образования для повышения эффективности образовательного процесса.

ПОСТАНОВКА ЗАДАЧИ

Цель и задачи: разработать информационную систему (ИС) для многофакторного анализа учебной деятельности образовательной организации с использованием технологий BigData.

Функции информационной системы.

1. Анализ абитуриентов, которые были зачислены в период 2014–2020 гг. по регионам России.

2. Анализ абитуриентов, которые были зачислены в период 2014–2020 гг. по странам мира.

3. Анализ количества отчисленных студентов, которые были зачислены из других регионов, в том числе по собственному желанию: по регионам, структурам и направлениям подготовки.

4. Процент отчисленных по собственному желанию по структурам и направлениям подготовки, в том числе в течение 1 курса.

5. Процент отчисленных студентов по собственному желанию после первого курса по структурам и направлениям подготовки, в том числе с баллом ЕГЭ ≥ 70 баллов; от 60 до 70 баллов; ≤ 60 баллов, в том числе после СПО.

6. Процент отчисленных за академическую неуспеваемость по структурам и направлениям подготовки, в том числе обучающихся за счет бюджета и внебюджета, в том числе после СПО.

7. Процент отчисленных студентов, поступивших по целевому договору, в том числе по собственному желанию: по структурам и направлениям подготовки.

8. Процент студентов, которым предоставляется академический отпуск: по структурам и направлениям подготовки.

9. Анализ успеваемости студентов, обучающихся на бюджетных и внебюджетных местах.

10. Анализ успеваемости студентов, сдавших ЕГЭ на ≥ 70 баллов; от 60 до 70 баллов; ≤ 60 баллов: по структурам и направлениям подготовки.

11. Анализ успеваемости студентов, поступивших после СПО, в том числе на бюджет и внебюджет.

12. Анализ успеваемости студентов, поступивших в рамках особой квоты: по структурам и направлениям подготовки, в том числе после СПО.

13. Сравнение успеваемости в бакалавриате и магистратуре (одного и того же человека).

14. Процент получивших государственную академическую стипендию в повышенном размере (3–4 курса) среди тех, кто приехал из других регионов: по структурам и направлениям подготовки.

15. Процентное соотношение студентов, получающих государственную академическую стипендию в повышенном размере, поступивших после школы и СПО: по структурам и направлениям подготовки.

16. Баллы ЕГЭ студентов, получающих государственную академическую стипендию в повышенном размере.

Для выполнения данных функций в информационной системе реализованы алгоритмы хранения и обработки данных о студентах, обучающихся в Тамбовском государственном университете им. Г.Р. Державина, всех курсах и образовательных программах, а также подробная информация о каждой из образовательных программ университета.

Для заполнения базы данных была использована сгенерированная тестовая выборка данных с 2014 по 2020 г.

РЕАЛИЗАЦИЯ

Для написания базовых функций ядра информационной системы использовался язык высокого уровня Python, позволяющий быстро и надежно разрабатывать программы для анализа информации.

Список библиотек Python, которые использовались в ядре системы:

1) Pandas – инструмент, предназначенный для быстрой и простой обработки данных, чтения, агрегирования и визуализации;

2) NumPy – предоставляет высокопроизводительные объекты многомерных массивов и инструменты для работы с ними;

3) SciPy – содержит модули для эффективных математических процедур, таких как линейная алгебра, интерполяция, оптимизация, интеграция и статистика¹;

4) Matplotlib, Seaborn – библиотеки для визуализации, построения графиков проведенного анализа;

5) Для реализации графического интерфейса программного средства использовались в связке PyQt5 и QtDesigner;

6) PyQt5 – наиболее часто используемый модуль для создания GUI приложений в Python;

7) QtDesigner – это инструмент для проектирования и построения графических пользовательских интерфейсов (GUI) с помощью наборов элементов для создания классических пользовательских интерфейсов в

¹ Топ-10 библиотек Python для Data Science. URL: <https://datastart.ru/blog/read/top-10-bibliotek-python-dlya-data-science> (дата обращения: 26.02.2021).

настольном стиле, позволяющий простым перетаскиванием построить макет готового приложения и органично интегрировать его с запрограммированным кодом, используя механизм сигналов и слотов.

Для реализации всех функций информационной системы была спроектирована структура и архитектура планируемой базы данных. Данная структура представлена на рис. 1.

Для администрирования базы данных используется СУБД MySQL и инструмент, предназначенный для упрощения администрирования созданной базы данных, – phpMyAdmin.

MySQL – база данных, использующая универсальный язык SQL, которая представляет собой программу, помогающую хранить информацию и обеспечивать функциональность для создания, модификации и управления данными в реляционных базах данных, и делает это быстро. Она хранит каждый вид данных в виде простых таблиц, разбитых на строки и столбцы, на пересечении которых расположены данные. Запросы к таким таблицам возвращают таблицы, которые сами могут становиться предметом дальнейших запросов. Каждая база данных может включать несколько таблиц, которые, как правило, связаны друг с другом, откуда и произошло название реляционные [2].

Для написания кода программы используется интегрированная среда разработки PyCharmCommunity, IDE для языка программирования Python, которая предоставляет средства для анализа кода, графический отладчик, инструмент для запуска юнит-тестов.

На рис. 2 представлены результаты анализа данных по студентам, которые были зачислены в университет из различных регионов нашей страны.

На рис. 3 представлены результаты анализа данных по студентам, которые были зачислены в университет из различных стран мира.

ВЫВОДЫ

В настоящее время спроектирована структура данных, реализована модель данных с MySQL, разработан графический интерфейс пользователя. Произведена генерация значений, соответствующих определенным требованиям. В дальнейшем планируется обучить модель на реальных данных и расширить функционал ИС, а также интегрировать данную систему в цифровую среду университета.

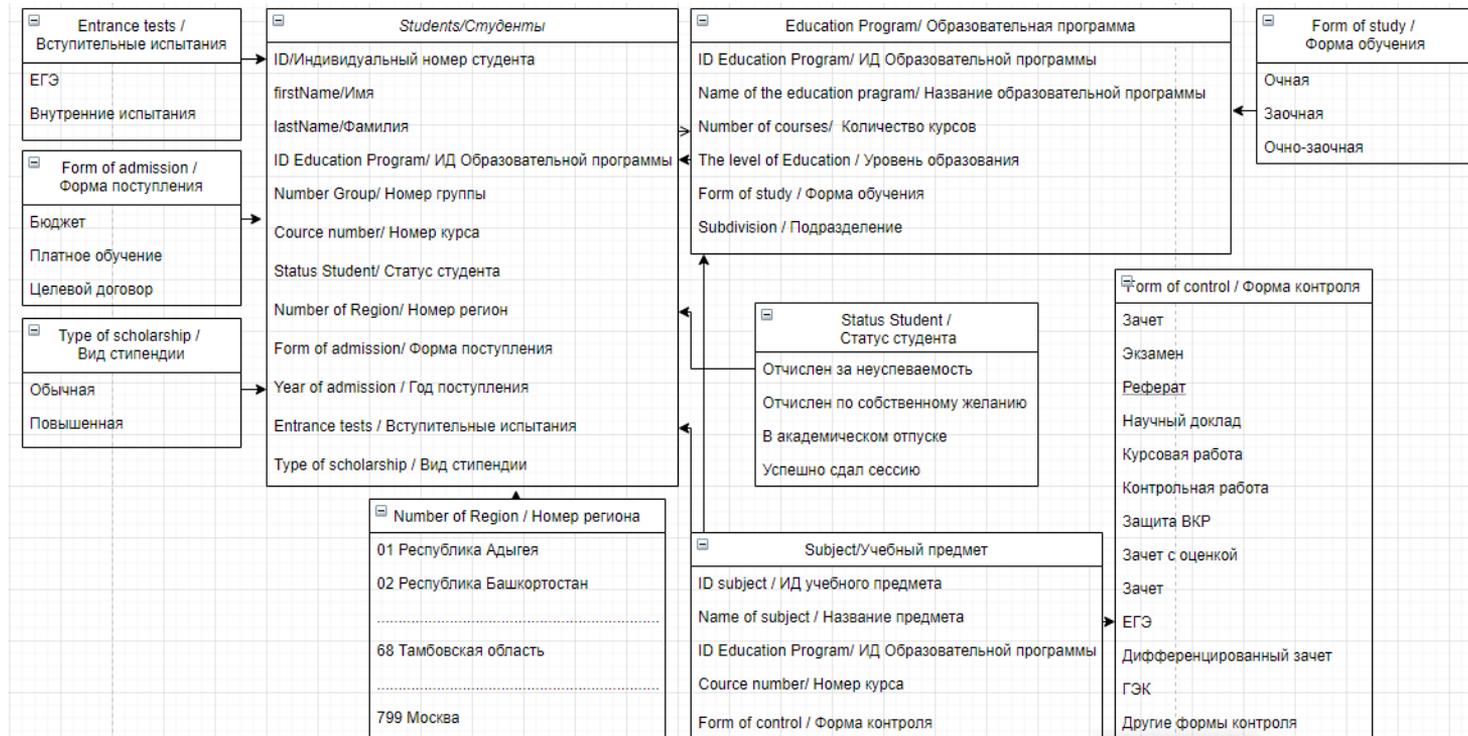


Рис. 1. Структура данных ИС

Анализ								
Анализ абитуриентов, которые были зачислены в период 2014 – 2020 годы по регионам								
Регион	2014	▼	2015	2016	2017	2018	2019	2020
Тамбовская область	733		798	786	743	750	713	707
Саратовская область	19		7	11	0	3	8	4
Рязанская область	15		12	14	10	15	18	14
Воронежская область	11		17	9	14	5	12	18
Липецкая область	9		8	5	8	8	13	10
Брянская область	2		1	0	2	2	1	0
Волгоградская область	2		0	1	0	1	2	2
Ивановская область	2		1	1	2	0	0	0
Иркутская область	2		1	1	1	0	0	0
Калужская область	2		2	1	1	2	0	1
Карачаево-Черкесская Республика	2		2	2	2	2	1	1
Кировская область	2		2	2	1	1	1	1
Костромская область	2		2	1	1	0	2	2
Краснодарский край	2		1	1	0	0	0	0
Курганская область	2		1	1	2	1	2	0
Нижегородская область	2		1	0	0	2	0	1
Новосибирская область	2		2	2	2	1	1	1
Оренбургская область	2		0	2	1	0	0	1
Пензенская область	2		0	0	0	2	0	2
Пермский край	2		1	1	1	2	2	1
Республика Алтай	2		2	0	0	0	2	2
Республика Тува	2		1	1	0	2	0	1
Ростовская область	2		0	2	0	2	0	2
Самарская область	2		1	1	2	1	1	2
Санкт-Петербург	2		1	2	2	1	1	1
Свердловская область	2		0	1	1	1	2	2
Ставропольский край	2		0	0	1	2	2	1
Тверская область	2		2	2	2	2	1	0

Рис. 2. Результаты анализа данных по регионам России, из которых были зачислены студенты

Анализ							
Анализ данных по студентам, которые были зачислены в университет из различных стран мира в период с 2014 до 2020 года							
Страна	2014	2015	2016	2017	2018	2019	2020
Россия	1043	1092	1161	1019	1067	1010	1124
Индия	37	34	30	35	36	38	36
Узбекистан	13	19	17	14	13	18	15
Китай	13	11	17	12	16	10	17
Казахстан	11	18	13	11	19	16	10
Сирия	9	9	7	9	6	7	6
Египет	5	8	6	8	7	9	8
Молдавия	4	2	3	1	3	2	2
Азербайджан	2	0	4	4	4	1	2
Киргизия	2	4	2	4	1	4	2
Гамбия	2	1	0	0	0	2	2
Ирак	2	1	1	1	0	0	1
Камерун	2	2	1	0	2	1	1
Ливан	2	1	2	2	2	1	2
Нигерия	2	1	2	0	1	1	2
Тунис	2	2	1	2	0	2	2
ЮАР	2	2	2	0	1	1	0
Бенин	1	2	2	0	1	2	0
Габон	1	0	0	2	0	0	2
Зимбабве	1	1	2	1	2	2	1
Либерия	1	0	1	2	2	2	2
Мали	1	2	0	1	0	0	0
Мозамбик	1	1	0	0	0	1	2
Оман	1	2	0	2	1	2	1
Палестина	1	0	2	0	2	1	2
Сенегал	1	1	2	2	2	1	2
Армения	0	1	4	0	2	4	2
Белоруссия	0	1	2	0	1	2	0

Рис. 3. Результаты анализа данных по странам, из которых были зачислены студенты

Список литературы

1. *Марц Н., Уоррен Дж.* Большие данные: принципы и практика построения масштабируемых систем обработки данных в реальном времени. Москва; Санкт-Петербург; Киев: Вильямс, 2016. 356 с.
2. *Гордеев С.И., Волошина В.Н.* Организация баз данных. Ч. 2. М.: Юрайт, 2019. 501 с.

Статья поступила в редакцию 09.03.2021
Одобрена после рецензирования 01.04.2021
Принята к публикации 06.04.2021

Информация об авторе:

Агапов Павел Викторович, магистрант по направлению подготовки «Математика», Тамбовский государственный университет им. Г.П. Державина, г. Тамбов, Российская Федерация, agapov.pasha.2012@gmail.com

INFORMATION SYSTEM FOR MULTIVARIATE ANALYSIS OF EDUCATIONAL ACTIVITIES OF AN EDUCATIONAL ORGANIZATION USING BIGDATA TECHNOLOGIES

Pavel V. Agapov, Master's Degree Student in "Mathematics" Programme, Derzhavin Tambov State University, Tambov, Russian Federation, agapov.pasha.2012@gmail.com

Abstract. We provide the main results of the development of a data structure and a software tool for the analysis of the educational process at the Derzhavin Tambov State University. Anonymized generated table values are used as input data. The main goals and objectives of the project, its functionality, as well as the implementation of an intelligent platform using the object-oriented programming language Python are described. In the future, we plan to expand the functionality of the developed software tool.

Keywords: data analysis, database, big data, software, Python, MySQL

The article was submitted 09.03.2021
Approved after reviewing 01.04.2021
Accepted for publication 06.04.2021