

ПОСТРОЕНИЕ ХРАНИЛИЩА ДАННЫХ ДЛЯ ПОСЛЕДУЮЩЕГО ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

В статье рассмотрены перечень данных, пригодных для анализа успеваемости студентов, а также выбранные таблицы на примере внутренней базы данных университета; выполнены построение запроса на основе выбранных таблиц, проверка полученного результата и анализ полученных результатов. Впоследствии полученные результаты будут использованы для построения хранилища данных и интеллектуального анализа.

Ключевые слова: хранилища данных (ХД), интеллектуальный анализ данных (ИАД), УРИС, балльно-рейтинговая система, универсальная рейтинговая информационная система, оценка качества образования вуза, оценка эффективности пары преподаватель – студент, экспериментальный образец универсальной информационной рейтинговой системы.

DATA WAREHOUSE AND DATA MINING

In the article, the list of data suitable for analysis of student performance, and the use selected tables on the example of the internal database of university are considered. We made the query, based on the selected tables. We check and investigate the result. In the future, the results will be used to build a data warehouse and data mining.

Keywords: data warehouse (DWH), data mining (DM), URIS, cumulative standing, universal rating information systems, education quality assessment of university, performance evaluation pair teacher – student, experimental model of universal information rating system.

Объектом исследования являются высшие учебные заведения (далее – вуз), предметом исследования является успеваемость студентов [2; 3; 4; 5], для исследования успеваемости необходимо определить набор данных, пригодных для анализа. Целью данной статьи является описание процесса создания хранилища (набора) данных для дальнейшего интеллектуального анализа данных.

В основу набора данных, пригодных для анализа, возьмем результаты, полученные Соболевым К.О. в работе [1]; из предложенного перечня возьмем интересующие нас атрибуты. Перечень атрибутов, необходимых для построения хранилища, приведен в таблице 1.

Таблица 1

Перечень атрибутов, необходимых для построения хранилища

№ п/п	Наименование атрибутов
1.	Наименование вуза

¹ Студент АНО ВО «Российский новый университет».

2.	Наименование факультета
3.	Наименование группы
4.	Регион вуза
5.	ФО вуза
6.	ФИО студента
7.	Год рождения студента
8.	Пол студента
9.	Наименование дисциплины
10.	Семестр изучения дисциплины
11.	Тип аттестации по дисциплине
12.	Общее время изучения дисциплины
13.	Общее время обучения по специальности в рамках учебного плана
14.	Количество часов всего по дисциплине
15.	Оценка по дисциплине

В качестве базы данных будет использована внутренняя база данных (далее – ВБД).

Задачи:

- поиск таблиц в ВБД, согласно перечню, перечисленному в таблице 1;
- создание запроса, формирующего необходимый перечень данных;
- проверка полученного результата.

ВБД состоит из 240 таблиц, описание структуры хранения и атрибутного состава не дано. Было проведено исследование всех таблиц на наличие интересующих нас атрибутов. Описанный выше перечень данных был обнаружен в следующих таблицах:

- tblScholasticPlanResult;
- tblOrderGoodsStudent;
- tblStudent;

- tblScholasticPlan;
- tblResult;
- tblGoods;
- tblExamination;
- tblCheckingType.

Пример взаимосвязи таблиц изображен на рис. 1.

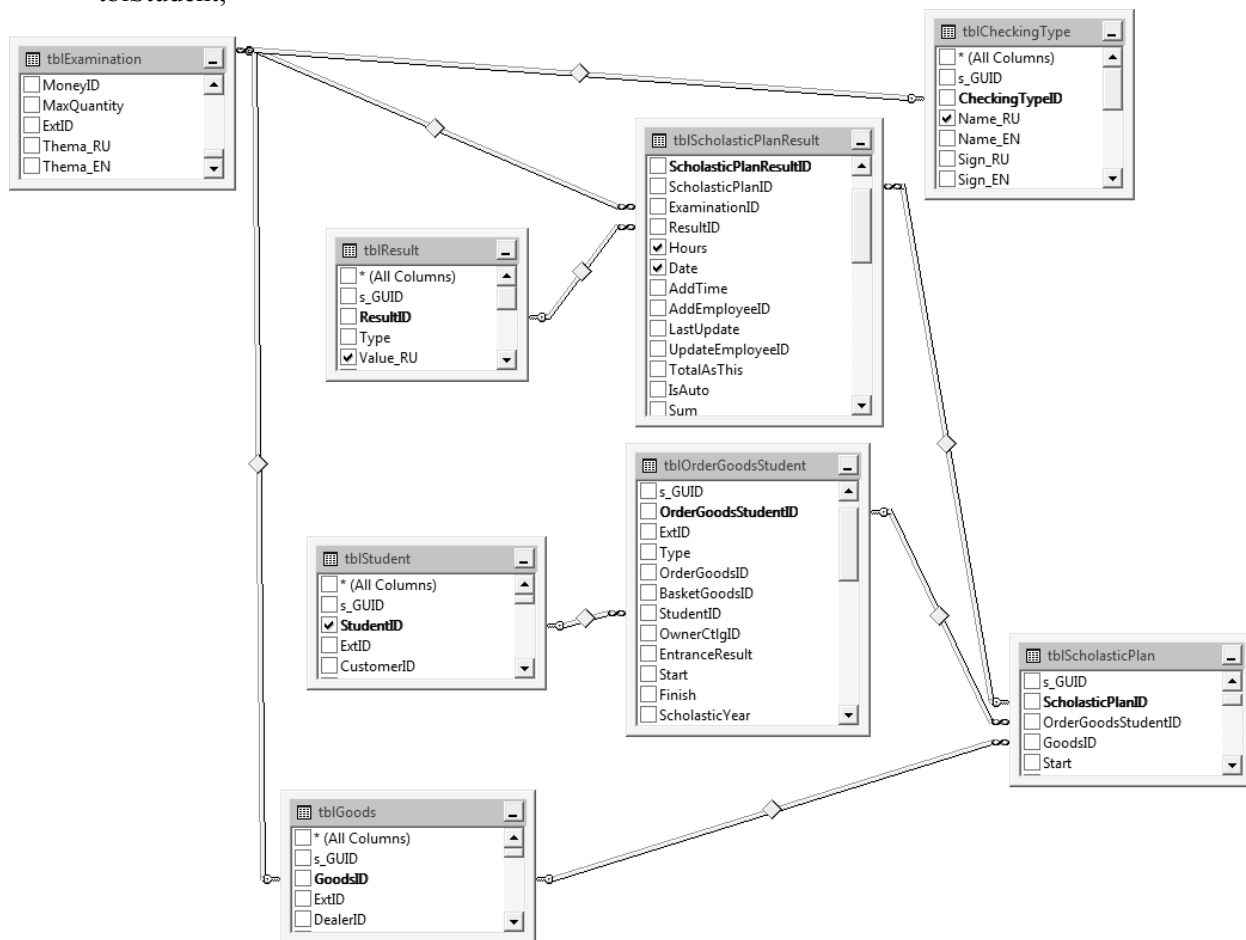


Рис. 1. Взаимосвязи таблиц, необходимых для получения данных

TBLSCHOLASTICPLANRESULT – в данной таблице содержится информация об учебных планах. Перечень данных, необходимый для создания вышеописанного набора данных (далее – набор данных), перечислен в табл. 2.

TBLSTUDENT – содержит данные о студенте. Перечень данных, используемых в наборе данных, перечислен в таблице 3.

TBLGOODS – содержится информация об оценках. Перечень данных, используемых в наборе данных, перечислен в таблице 4.

Таблица 2

Перечень необходимых данных из таблицы tblScholasticPlanResult

№ п/п	Наименование столбца	Тип данных	Наименование
1.	Hour	INT	Общее время по учебному плану
2.	Date	DATETIME	Дата сдачи предмета
3.	Semester	INT	Семестр изучения дисциплины

Таблица 3

**Перечень необходимых данных
из таблицы tblStudent**

№ п/п	Наименование столбца	Тип данных	Наименование
1.	First Name_RU	VAR-CHAR(50)	Имя студента
2.	Middle Name_RU	VAR-CHAR(50)	Отчество студента
3.	Last Name_RU	VAR-CHAR(50)	Фамилия студента
4.	Birthday	DATETIME	Дата рождения
5.	Sex	INT	Пол

Таблица 4

**Перечень необходимых данных
из таблицы tblGoods**

№ п/п	Наименование столбца	Тип данных	Наименование
1.	Name_RU	VAR-CHAR(150)	Название предмета

TBLORDERGOODSSTUDENT – содержит информацию о зачисленных студентах, будет применяться в дальнейшем для сокращения объема выборки с улучшением ее качества.

TBLRESULT – словарь результатов. Перечень данных, используемых в наборе данных/запросе, перечислен в таблице 5.

Таблица 5

**Перечень необходимых данных
из таблицы tblResult**

№ п/п	Наименование столбца	Тип данных	Наименование
1.	Value_RU	VAR-CHAR(50)	Оценка по предмету (словарь)
2.	Sign_RU	VAR-CHAR(10)	Сокращение оценки по предмету (словарь)

TBLCHECKINGTYPE – словарь испытаний по предмету. Перечень данных, используемых в наборе данных, перечислен в таблице 6.

TBLEXAMINATION – содержит информацию об испытаниях по предмету, будет применяться в дальнейшем для сокращения объема выборки с улучшением ее качества.

Таблица 6

**Перечень необходимых данных
из таблицы tblCheckingType**

№ п/п	Наименование столбца	Тип данных	Наименование
1.	Name_RU	VAR-CHAR(100)	Название испытания по предмету

TBLSCHOLASTICPLAN – содержит информацию об общем времени на обучение. Перечень данных, используемых в наборе данных, перечислен в таблице 7.

Таблица 7

**Перечень необходимых данных
из таблицы tblScholasticPla**

№ п/п	Наименование столбца	Тип данных	Наименование
1.	TotalHours	INT	Количество часов всего по дисциплине, в заданном семестре
2.	ClassHours	INT	Количество аудиторных часов
3.	HomeHour	INT	Количество часов для домашнего изучения

Итоговый набор данных, сопоставленный с найденными данными, перечислен в таблице 8.

Таблица 8

Итоговый набор данных

№ п/п	Наименование столбца	Тип хранения	Место хранения
1.	Наименование вуза	Расчетное	–
2.	Наименование факультета	Хранимое	tblOrder-GoodsStudent
3.	Наименование группы	Хранимое	tblOrder-GoodsStudent
4.	Регион вуза	Расчетное	–
5.	ФО вуза	Расчетное	–
6.	ФИО студента	Хранимое	tblStudent
7.	Год рождения студента	Хранимое	tblStudent

	Пол студента	Хранимое	tblStudent
	Наименование дисциплины	Хранимое	tblGoods
	Семестр изучения дисциплины	Хранимое	tblScholasticPlanResult
	Тип аттестации по дисциплине	Хранимое	tblCheckingType
	Общее время изучения дисциплины	Расчетное	tblScholasticPlan

	Общее время обучения по специальности в рамках учебного плана	Расчетное	tblScholasticPlan
	Количество часов всего по дисциплине	Расчетное	tblScholasticPlan
	Оценка по дисциплине	Хранимое	tblResult

Визуальный результат созданного запроса изображен на рис. 2.

ID_Студента	Фамилия	Имя	Отчество	Часы	Дата	Рейтинг	Предмет
21821	Волков	Антон	Игоревич	90	2005-09-15 00:00:00.000	1	Отечественна
23934	Бурцев	Виктор	Леонидович	90	2005-06-14 00:00:00.000	1	Менеджмент
24033	Симонов	Виктор	Алексеевич	90	2005-06-14 00:00:00.000	1	Менеджмент
24052	Степанько	Илья	Анатольевич	90	2005-06-14 00:00:00.000	1	Менеджмент
24432	Аллямшина	Альфия	Хамматовна	90	2005-01-21 00:00:00.000	1	Отечественна
24436	Гладуш	Ирина	Олеговна	90	2005-01-21 00:00:00.000	1	Отечественна
24440	Душин	Кирилл	Владимирович	90	2005-01-21 00:00:00.000	1	Отечественна
24444	Бушуев	Данила	Юрьевич	90	2005-01-21 00:00:00.000	1	Отечественна
24447	Голденкова	Дарья	Игоревна	90	2005-01-21 00:00:00.000	1	Отечественна
24456	Иванова	Екатерина	Александровна	90	2005-01-21 00:00:00.000	1	Отечественна
24459	Лидаков	Павел	Александрович	90	2005-01-21 00:00:00.000	1	Отечественна
24461	Замуреева	Алина	Сергеевна	90	2005-01-21 00:00:00.000	1	Отечественна
24464	Захарова	Ольга	Сергеевна	90	2005-01-21 00:00:00.000	1	Отечественна
24467	Куликова	Ольга	Павловна	90	2005-01-21 00:00:00.000	1	Отечественна
24471	Мещеряков	Анатолий	Викторович	90	2005-01-21 00:00:00.000	1	Отечественна
24473	Макунин	Александр	Николаевич	90	2005-01-21 00:00:00.000	1	Отечественна
24478	Нестерюк	Сергей	Валентинович	90	2005-01-21 00:00:00.000	1	Отечественна
24485	Намоконов	Евгений	Сергеевич	90	2005-01-21 00:00:00.000	1	Отечественна
24493	Ракова	Наталья	Игоревна	90	2005-01-21 00:00:00.000	1	Отечественна
24499	Спиридонова	Инна	Андреевна	90	2005-01-21 00:00:00.000	1	Отечественна
24503	Тыщенко	Алла	Николаевна	90	2005-01-21 00:00:00.000	1	Отечественна
24507	Ховренко	Екатерина	Андреевна	90	2005-01-21 00:00:00.000	1	Отечественна
24511	Хаустов	Артем	Сергеевич	90	2005-01-21 00:00:00.000	1	Отечественна
24515	Шкарупа	Алексей	Витальевич	90	2005-01-21 00:00:00.000	1	Отечественна
24518	Щедурских	Дмитрий	Александрович	90	2005-01-21 00:00:00.000	1	Отечественна

Рис. 2. Результат работы запроса

Результат запроса будет использован для дальнейшего построения хранилища данных. При работе с полученными данными определено более 700 000 записей по 22 000 субъектов исследования.

Выводы

В статье предложена реализация ненормализованного хранилища данных для последующего интеллектуального анализа данных. В основу хранилища легла ВБД, содержащая порядка 700 000 записей по 22 000 субъектов. Полнота данных достаточна для дальнейшего изучения и применения методов кластеризации и классифи-

кации и иных методов интеллектуального анализа данных.

Литература

1. Соболев К.О. Хранилище данных универсальной рейтинговой информационной системы // Т-Сотм: Телекоммуникации и транспорт. – 2015. – Т. 9. – № 2. – С. 74–78.
2. Соболев К.О., Брюсова В.О. Анализ студенческой успеваемости на примере кафедры информационных технологий и естественнонаучных дисциплин // Решение – 2014 : материалы Третьей Всероссийской научно-практической

конференции школьников и студентов. – Пермь : Березниковский филиал Пермского национального исследовательского политехнического университета, 2014. – С. 101–103.

3. Широков О.Н. Актуальные направления развития научной и образовательной деятельности // Сб. науч. тр. (Чебоксары, 22 апр. 2014 г.). – Чебоксары : ЦНС «Интерактив плюс», 2014. – 295 с.

4. Степанов А.Г., Бабенков А.Н., Блюм В.С. и др. Информационные технологии в работе кафедры. – СПб., 2014.

5. Крылова А.Г. Успеваемость студентов экономических специальностей и статистическая оценка факторов, на нее влияющих // Современные научные исследования и инновации. – 2012. – № 6 [Электронный ресурс]. – Режим доступа: <http://web.snauka.ru/issues/2012/06/14688>.