

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»

И.Ф. Астахова, В.А. Чулюков, И.П. Половинкин

ПРОЕКТИРОВАНИЕ БАЗ ДАННЫХ

Учебное пособие

Воронеж
Издательский дом ВГУ
2017

ВВЕДЕНИЕ

Информационные системы с базами данных являются в настоящее время одной из важнейших областей современных компьютерных технологий. С этой сферой связана большая часть современного рынка программных продуктов. Одной из общих тенденций в развитии систем с базами данных являются процессы интеграции и стандартизации, затрагивающие структуры данных и способы их обработки и интерпретации, системное и прикладное программное обеспечение, средства разработки систем взаимодействия их компонентов и т.д. Основой современных систем управления базами данных (СУБД) является реляционная модель представления данных – в большой степени благодаря простоте и четкости ее концептуальных понятий и строгого математического обоснования.

Данное пособие предназначено в первую очередь для преподавателей и студентов и ориентировано на обучение основам проектирования информационных систем с базами данных.

Авторы надеются, что пособие окажется полезным не только для преподавателей и студентов, но и для других читателей, заинтересованных в получении начальных навыков проектирования и использования информационных систем с базами данных.

ванная информация является избыточной. Если, например, удалить один из номеров кафедры информатики, эта информация может быть получена из других кортежей отношения. На рис. 2 (б) показано, как будет выглядеть отношение в случае замещения дублированных телефонных номеров «пропущенными» значениями (или так называемыми null-значениями).

С-К-Т		
СЛЖ#	КАФ	ТЕЛ#
102	Информатики	95
127	Физики	34
187	Физики	34
230	Информатики	95

a

С-К-Т		
СЛЖ#	КАФ	ТЕЛ#
102	Информатики	95
127	Физики	34
187	Физики	-
230	Информатики	-

б

С-К	
СЛЖ#	КАФ
102	Информатики
127	Физики
187	Физики
230	Информатики

К-Т	
КАФ	ТЕЛ#
Информатики	95
Физики	34

в

Рис. 2. Исключение избыточных данных

Конечно, телефонные номера кафедр Физики и Информатики могут быть найдены из других кортежей, т.е. данные не утеряны. Однако, необходимы дополнительные меры, чтобы узнать реальные значения «нулей», например, номер телефона служащего с табельным номером 187. Более того, отношение, представленное на рис. 2 (б), имеет структуру, которая может нарушить целостность данных при удалении информации. Если служащий с СЛЖ# = 127 увольняется, кортеж <127, Физики, 34> должен быть удален из отношения. Следовательно, произойдет утеря телефонного номера кафедры Физики, поскольку нигде больше в отношении он не представлен. Таким образом, рассмотренный метод управления избыточностью неудовлетворителен.

На рис. 2 (в) показан лучший способ исключения избыточности телефонных номеров. Отношение С-К-Т заменено двумя отношениями. От-

ношение С-К содержит информацию о табельных номерах служащих и месте их работы, а отношение К-Т – о телефонных номерах кафедр. Теперь при удалении информации о служащем 127 информация о номере телефона кафедры Физики утеряна не будет.

Задача 3: Сведение числа хранимых в БД отношений к минимуму.

Разбиение одного отношения на два или более меньших отношений желательно для исключения ряда проблем. Но это неудобно для пользователя. Таким образом, нельзя допускать неограниченный рост числа отношений.

Задача 4: Нормализация отношений.

Для некоторых отношений очень важны проблемы удаления, обновления и вставки. Поэтому потенциально опасные отношения должны быть нормализованы. *Нормализация* – это разбиение одного отношения на два или более в соответствии со специальной процедурой разбиения.

1.2. Универсальное отношение

Рассмотрим несколько модифицированный пример, взятый из [3]. Пусть требуется разработать небольшую БД для деканата факультета. На первом шаге проектирования необходимо определить все атрибуты, которые необходимы деканату в БД, и связи между атрибутами:

- Сном: Номер зачетной книжки студента. Целое значение, уникальное для каждого студента университета.
- Сфам: Фамилия студента. Каждый студент имеет только одну фамилию, но не исключено, что одну фамилию носят несколько студентов.
- Кном: Номер комнаты в общежитии. Каждый студент живет в общежитии и имеет комнату. В одной комнате может проживать более одного студента.
- Тном: Номер телефона студента. Каждая комната общежития имеет один внутренний телефон, и им пользуются все студенты, проживающие в этой комнате.

- Курс:** Идентификационный номер курса, пройденный студентом. Например, МТА – это идентификационный номер курса математического анализа.
- Семестр:** Идентификационный номер семестра, в котором данный курс был завершен студентом. Например, О99 означает осенний семестр 1999 г., а 398 – зимний семестр 1998 года. Возможно, что студент изучал один и тот же курс в различных семестрах.
- Оценка:** Оценка за курс. Оценка, полученная студентом за определенный курс в данном семестре.

На рис. 3 показан образец данных, представленных деканатом для их хранения в БД.

ФАКУЛЬТЕТ

Сном	Сфам	Кном	Тном	Курс	Семестр	Оценка
3297	Иванов	120	136	МТА	О97	4
				ИНФ	О97	5
				ФИЗ	398	2
				МТА	398	4
3496	Петров	238	344	МТА	О96	3
				МТА	397	4
				ПСИ	398	5
3596	Сидоров	120	136	АЛГ	397	2
				ГЕО	О97	4
				ФИЗ	О97	3
4798	Николаев	345	321	ПЕД	399	4

Рис. 3. Данные, необходимые деканату

Указанная таблица отношением не является. Для иллюстрации этого выделим одну «строку» из таблицы (рис. 3). На этом рисунке значения четырех полей Сном, Сфам, Кном и Тном – атомарные. *Атомарным* (или *скалярным*) называется неделимое значение, а не множество, или кортеж значений из некоторых доменов. Другими словами, в каждом отношении в позиции на пересечении столбца и строки всегда находится только одно

значение и никак не группа из нескольких значений. В нашем случае (рис. 3) значения в полях Курс, Семестр и Оценка множественные. Эти поля – пример того, что называют **группой повторения** (рис. 4). Группа повторения – это поле или комбинация полей, которая содержит несколько значений данных в каждом кортеже (в общем случае разное количество значений в разных кортежах). *В реляционных базах данных группы повторения недопустимы.*

Сном	Сфам	Кном	Тном	Курс	Семестр	Оценка
3297	Иванов	120	136	МТА	О97	4
				ИНФ	О97	5
				ФИЗ	398	2
				МТА	398	4

Рис. 4. Строка с группой повторения

Для того, чтобы таблица, приведенная на рис. 3 могла считаться отношением реляционной базы данных, ее надо преобразовать так, чтобы она не содержала групп повторения, т.е., чтобы каждый элемент кортежа имел атомарное значение. Обычно для этого используется простая вставка (рис. 5).

ФАКУЛЬТЕТ

Сном	Сфам	Кном	Тном	Курс	Семестр	Оценка
3297	Иванов	120	136	МТА	О97	4
3297	Иванов	120	136	ИНФ	О97	5
3297	Иванов	120	136	ФИЗ	398	2
3297	Иванов	120	136	МТА	398	4
3496	Петров	238	344	МТА	О96	3
3496	Петров	238	344	МТА	397	4
3496	Петров	238	344	ПСИ	398	5
3596	Сидоров	120	136	АЛГ	397	2
3596	Сидоров	120	136	ГЕО	О97	4
3596	Сидоров	120	136	ФИЗ	О97	3
4798	Николаев	345	321	ПЕД	399	4

Рис. 5. Корректное отношение

В результате в отношении появляется большой объем избыточных данных. Но таблица на рис. 5 представляет собой экземпляр *корректного* отношения. Его называют *универсальным отношением* проектируемой БД. В универсальное отношение включаются все представляющие интерес атрибуты, и оно может содержать все данные, которые предполагается размещать в БД. Универсальное отношение может служить отправной точкой при проектировании БД.

1.3. Аномалии единственного отношения

Чаще всего единственное отношение в БД не является лучшим способом хранения данных. Это обусловлено тем, что при определенных операциях на данные будет оказываться определенное воздействие, что может привести к следующим трем проблемам:

- 1) проблема, связанная с обновлением (изменением) данных в кортежах;
- 2) проблема, вызываемая необходимостью удаления кортежей;
- 3) проблема, обусловленная необходимостью добавления новых кортежей.

Эти проблемы называют аномалиями обновления, удаления и вставки.

Аномалия вставки

Для студента первого курса, еще не сдавшего ни один экзамен, необходимо включить в БД кортеж с null-значениями атрибутов Курс, Семестр и Оценка. Как уже отмечалось, в первичном ключе не может быть null-значений. Следовательно, включение в базу данных студентов первого курса в осеннем семестре невозможно вплоть до сдачи ими первого экзамена.

На рис. 6 показано отношение ФАКУЛЬТЕТ в случае включения в него информации о студенте Зайцеве, не прошедшем еще ни одного курса.

В число таких студентов попал Зайцев, хотя он не сдавал ни одного экзамена.