

КОНСПЕКТ ЗАНЯТИЯ № 1 ТРЕТЬЕЙ НЕДЕЛИ КУРСА «БАЗЫ ДАННЫХ»

1. СПЕЦИАЛЬНЫЕ ОПЕРАЦИИ РЕЛЯЦИОННОЙ АЛГЕБРЫ

Вспомним, какие операции называют специальными:

- ✓ проекция отношения;
- ✓ выборка отношения;
- ✓ соединение отношений;
- ✓ деление отношений.

Поговорим о каждой операции в отдельности.

Операция *проекции* является **унарной**, т.е. производится над одним отношением, и позволяет получить его вертикальный срез в виде интересующих нас атрибутов.

Более формально, *проекцией* отношения F по подмножеству его атрибутов $\{X, Y, \dots, Z\}$ называется отношение R ,

- схема которого состоит из подмножества атрибутов, по которым производилась проекция: $\{X, Y, \dots, Z\}$,
- а тело состоит из всех неповторяющихся кортежей $\{X:x, Y:y, \dots, Z:z\}$, таких, что в отношении F присутствует кортеж со значением x атрибута X , y атрибута Y , ... и z атрибута Z .

Варианты синтаксиса:

$$R = \Pi_{X, Y, \dots, Z}(F)$$

$$R = F[X, Y, \dots, Z]$$

Степень результирующего отношения равна количеству атрибутов, по которым производилась проекция.

Если в результате взятия проекции не возникнет дубликатов кортежей, кардинальность результата будет равна кардинальности исходного отношения (рис. 4.11).

Отношение F:
степень = 3, кардинальность = 3

X	Y	Z
x ₁	y ₁	z ₁
x ₂	y ₂	z ₂
x ₃	y ₃	z ₃

Отношение R = $\Pi_Y(F)$:
степень = 1, кардинальность = 3
(дубликатов не возникло)

Y
y ₁
y ₂
y ₃

Рис. 4.11. Проекция отношения, в результате которой не возникло дубликатов кортежей

В противном случае - будет меньше, т.к. для сохранения свойства уникальности кортежей отношения, возникшие дубли необходимо будет удалить (рис. 4.12).

Отношение F:
степень = 3, кардинальность = 3

X	Y	Z
x ₁	a ₁	z ₁
x ₂	y ₂	z ₂
x ₃	a ₁	z ₃

Отношение R = $\Pi_Y(F)$:
степень = 1, кардинальность = 3-1 = 2
(появились дубликаты)

Y
a ₁
y ₂
a ₁

Рис. 4.12. Проекция отношения, в результате которой возникли дубликаты кортежей

В качестве примера рассмотрим часть меню столовой, в котором для каждого блюда определено несколько вариантов порций, разного веса и стоимости. В результате взятия проекции этого отношения по атрибуту Название блюда, мы получим ассортимент всех блюд столовой, без учёта ненужных нам в этой ситуации сведений о весе и цене порции (рис. 4.13).

Отношение Меню

Название блюда	Вес порции	Цена
Рис	150	15
Рис	250	20
Картофель	150	25
Картофель	300	45

Отношение R = $\Pi_{\text{Название блюда}}(\text{Меню})$

Название блюда
Рис
Картофель

Рис. 4.13. Пример проекции отношения «Меню» по атрибуту «Название блюда»

Операция *выборки* также является унарной. В результате этой операции мы получаем горизонтальный срез отношения. Более формально, **выборкой или ограничением** отношения F по θ -предикату называется отношение R

- с той же схемой, что и у отношения F
- и телом, состоящим из подмножества кортежей отношения F, значения атрибутов которых удовлетворяют заданному предикату.

Варианты синтаксиса:

$$R = \sigma_{\theta\text{-предикат}}(F)$$

$$R = F [\theta\text{-предикат}] \text{ или } R = F \text{ WHERE } \theta\text{-предикат}$$

Операция выборки не влияет на количество атрибутов в отношении и поэтому степень результата останется неизменной.

Условия выборки

Простое условие выборки будем называть тета-предикатом (θ -предикатом).

θ -предикат соответствует сравнению значений двух атрибутов отношения (например, Зарплата > Премия) или значения атрибута и некоторой константы (например, Зарплата = 2000).

Условие выборки может быть и сложным, состоящим из простых тета-предикатов, объединённых логическими операторами.

Сложные условия выборки всегда могут быть обратно разбиты на простые:

- логический оператор AND аналогичен операции пересечения, т.е. нахождению множества кортежей, удовлетворяющим обоим простым предикатам одновременно:

$$\sigma_{\theta\text{-предикат1 AND } \theta\text{-предикат2}}(F) = \sigma_{\theta\text{-предикат1}}(F) \cap \sigma_{\theta\text{-предикат2}}(F)$$

- логическое отрицание производится путём удаления из первоначального отношения кортежей, удовлетворяющих исходному условию выборки:

$$\sigma_{\text{NOT } \theta\text{-предикат1}}(F) = F \setminus (\sigma_{\theta\text{-предикат1}}(F))$$

Допустим, нам нужно из меню выбрать информацию только о небольших порциях весом 150 грамм, а также о любых других, чья цена не слишком велика (не более 20 рублей). Соответствующий сложный

предикат для выборки и результат операции представлены на рисунке 4.14.

Отношение Меню			Отношение $R = \sigma_{\text{количество} = 150 \text{ OR } \text{цена} \leq 20}(\text{Меню})$		
Название блюда	Вес порции	Цена	Название блюда	Вес порции	Цена
Рис	150	15	Рис	150	15
Рис	250	20	Рис	250	20
Картофель	150	25	Картофель	150	25
Картофель	300	45			

Рис. 4.14. Пример выборки определённых позиций из отношения «Меню»

Вы можете самостоятельно убедиться, что операция объединения двух полученных по отдельности множеств кортежей (одного - со значением атрибута количество - 150, и второго - со значением атрибута цена - меньше или равно 20) даст тот же самый результат.

Операция *соединение* можно подразделить на два вида:

1. θ -соединение (соединение по условию);
2. Естественное или натуральное соединение.

θ -соединением (соединением по условию) двух отношений:

F, состоящего из атрибутов $\{F1, F2, \dots, Fn\}$ и S, состоящего из атрибутов $\{S1, S2, \dots, Sm\}$ по θ -предикату, называется отношение R,

- схема которого является результатом сцепления схем исходных отношений: $\{F1, F2, \dots, Fn, S1, S2, \dots, Sm\}$,
- а тело содержит такие кортежи из результата декартова произведения F и S, значения которых удовлетворяют заданному предикату.

Варианты синтаксиса:

$$R = F \bowtie_{F.X \theta S.Y} S$$

$$R = F [F.X \theta S.Y] S$$

Исходя из определения, кардинальность результата будет, как правило, меньше кардинальности декартова произведения исходных отношений (т.к. из него мы отбираем только часть кортежей, удовлетворяющих заданному условию), а степень результата останется

равной степени декартова произведения, т.е. сумме степеней исходных отношений (рис. 4.15).

Одноимённые атрибуты исходных отношений превращаем в разноимённые путём использования их полных названий.

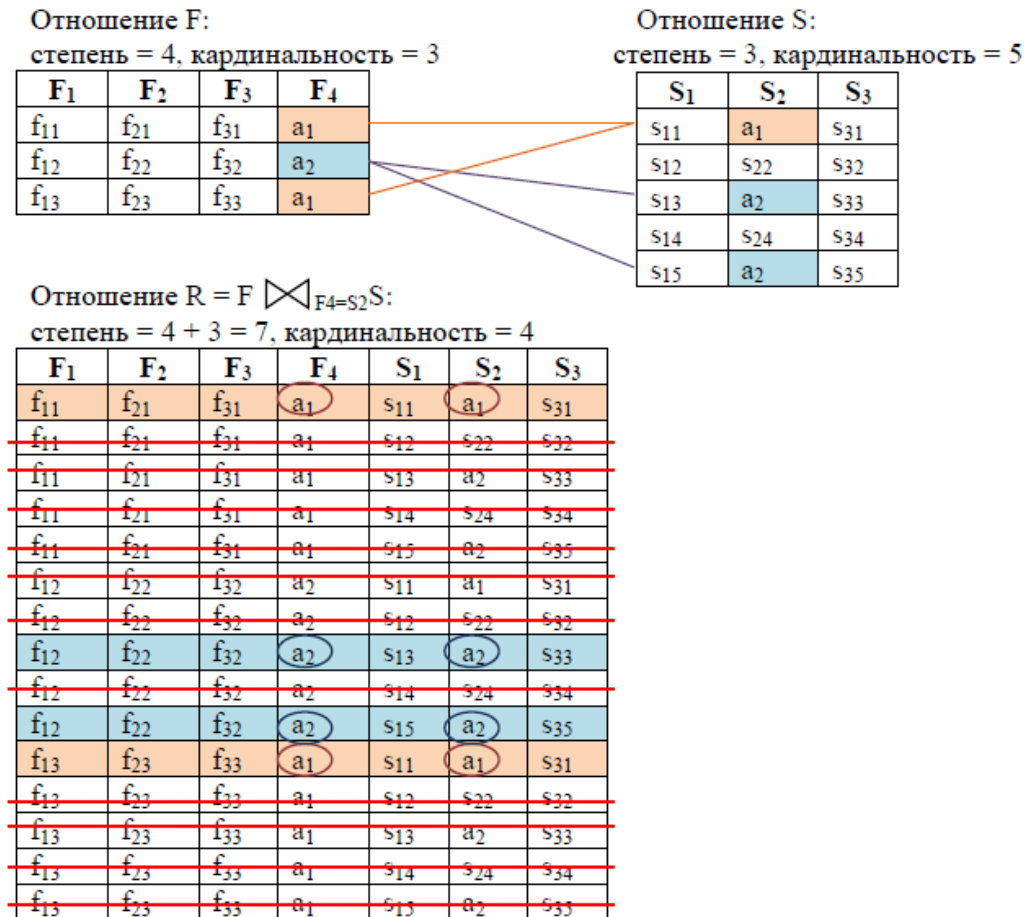


Рис. 4.15. θ -соединение отношений в общем виде

Условие соединения всегда должно включать сравнение между собой значений атрибутов соединяемых отношений.

Данная операция может быть выражена через другие операции реляционной алгебры (декартова произведения и выборки):

$$F \bowtie_{F.X \theta S.Y} S = (F \text{ TIMES } S) \text{ WHERE } X \theta Y$$

Естественное или натуральное соединение это частный случай, который был выделен отдельно за счет частого использования в решении профессиональных задач.

Естественное соединение представляет собой соединение отношений по условию равенства значений всех их общих атрибутов, т.е. атрибутов с одинаковыми именами, определённых на одинаковых доменах.

Естественным соединением двух отношений:

F, состоящего из атрибутов {X, Y}, возможно, составных, и S, состоящего из атрибутов {Y, Z}, возможно, составных, где Y – общий атрибут (группа атрибутов) исходных отношений, называется отношение R,

- схема которого состоит из подмножеств атрибутов {X, Y, Z},
- а тело содержит все такие кортежи {X:x, Y:y, ..., Z:z}, любой из которых присутствует и в отношении F, со значением x атрибута X и значением y атрибута Y, и в отношении S, со значением y атрибута Y и значением z атрибута Z.

Т.к. условие естественного соединения всегда одно и то же, его не нужно указывать при выполнении этой операции.

В качестве примера рассмотрим два вышеприведённых варианта соединения в рамках одной задачи. Пусть даны два отношения, одно из которых содержит сведения о сотрудниках организации, а второе – данные об отделах, в которых работают эти сотрудники (рис. 4.16). Требуется для каждого сотрудника отобразить полную информацию об отделе, в котором он работает.

Отношение Сотрудник				Отношение Отдел	
Номер договора	Фамилия	Оклад	Номер отдела	Номер отдела	Название отдела
10001	Васильчиков	50000	1	1	Маркетинг
10002	Краснин	65000	2	2	Программные технологии
10003	Золотарев	58500	2		

Рис. 4.16. Состояние исходных отношений «Сотрудник» и «Отдел»

Если явно задать условие соединения как равенство значений атрибутов номер отдела в обоих отношениях, то операция будет являться θ -соединением, и в результат должны будут попасть оба атрибута «Номер отдела», несмотря на то, что они содержат одинаковые данные.

Если провести естественное соединение – без явного указания условия – результат будет тем же за исключением наличия в нём только одного экземпляра общего для двух отношений атрибута Номер отдела (рис. 4.17).

Отношение R = Сотрудник ⋈_{Сотрудник.Номер отдела = Отдел.Номер отдела} Отдел

Номер договора	Фамилия	Оклад	Номер отдела	Номер отдела	Название отдела
10001	Васильчиков	50000	1	1	Маркетинг
10002	Краснин	65000	2	2	Программные технологии
10003	Золотарев	58500	2	2	Программные технологии

Отношение R = Сотрудник ⋈_{Отдел} Отдел

Номер договора	Фамилия	Оклад	Номер отдела	Название отдела
10001	Васильчиков	50000	1	Маркетинг
10002	Краснин	65000	2	Программные технологии
10003	Золотарев	58500	2	Программные технологии

Рис. 4.16. Сравнение естественного (внизу) и θ -соединения (наверху) отношений «Сотрудник» и «Отдел» по совпадению значений атрибута «Номер отдела»

Важно: в отличие от обычного θ -соединения, в результат которого должны попасть все без исключения атрибуты исходных отношений, в результате естественного соединения остаётся только по одному экземпляру каждого из общих атрибутов.

Важно: если у отношений нет общих атрибутов, то их естественное соединение будет равнозначно декартову произведению.

И наконец, последняя специальная операция в реляционной алгебре – *деление отношений*.

Пусть у нас есть два отношения:

- F, состоящее из атрибутов {X, Y}, возможно, составных и
- S, состоящее из атрибута {Y}, возможно, составного

Отношение F является делимым, его схема включает все без исключения атрибуты отношения-делителя S и, кроме того, содержит некоторое количество собственных атрибутов.

Результатом деления отношения F на отношение S будет являться некоторое отношение R:

- схема которого включает только собственные атрибуты отношения-делимого F: {X},

- а тело состоит из таких кортежей {X:x}, таких, что для любого кортежа {Y:y} $\in S$ в отношении F найдется кортеж {X:x, Y:y}

Т.е. декартово произведение каждого из кортежей отношения-результата R с отношением-делителем S полностью входит в состав отношения-делимого F (рис. 4.17).

Варианты синтаксиса:

$$R = F \div S$$

$$R = F \text{ DIVIDE BY } S$$

Кардинальность результата деления должна быть меньше или равна частному от деления кардинальностей исходных отношений. Степень результата всегда равна разности степеней исходных отношений.

Отношение F:

степень = 4, кардинальность = 3

X ₁	X ₂	Y ₁	Y ₂
x ₁₁	x ₂₁	a ₁	a ₂
x ₁₂	x ₂₂	a ₁	a ₂
x ₁₂	x ₂₂	b ₁	b ₂

Отношение S:

степень = 2, кардинальность = 2

Y ₁	Y ₂
a ₁	a ₂
b ₁	b ₂

Отношение R ÷ S:

степень = 4 - 2 = 2, кардинальность = 1 (3 ÷ 2 без дробной части)

X ₁	X ₂
x ₁₂	x ₂₂

Отношение Π_{X₁, X₂}(F) × S

(декартово произведение кортежей, состоящих только из собственных атрибутов отношения F с отношением S)

X ₁	X ₂	S.Y ₁	S.Y ₂
x ₁₁	x ₂₁	a ₁	a ₂
x ₁₁	x ₂₁	b ₁	b ₂
x ₁₂	x ₂₂	a ₁	a ₂
x ₁₂	x ₂₂	b ₁	b ₂

← этот кортеж отсутствует в отношении F, поэтому сочетание (x₁₁, x₂₁) не попадёт в результат деления

Рис. 4.17. Деление отношений в общем виде

Деление является ещё одной избыточной операцией, которая может быть выражена через сочетание операций декартова произведения, разности и выборки. В данном случае это большой плюс, т.к. в языке SQL деление как отдельная операция не реализовано.

Важно: распознать, что для получения нужного результата требуется выполнить операцию деления можно по слову “все” в формулировке условия задачи.

Рассмотрим следующий пример: пусть у нас есть отношение с информацией о положительных оценках студентов за экзамены и

отношение со списком предметов, по которым студентам нужно сдать экзамены в этом семестре (рис. 4.18). Требуется найти список студентов, успешно сдавших экзамены по всем предметам.

Отношение Вedomость				Отношение Предмет	
Номер зачётной книжки	Фамилия	Оценка	Предмет	Предмет	
123456	Савицкий	3	Физика	Физика	
123457	Иванова	4	Математика	Математика	
123458	Паулайнен	5	Математика	Русский язык	
123456	Савицкий	4	Математика		
123458	Паулайнен	4	Русский язык	Русский язык	
123456	Савицкий	5	Русский язык		

Рис. 4.18. Состояние исходных отношений «Вedomость» и «Предмет»

Если мы возьмёмся разделить отношение F в его исходном виде на отношение S в результат не попадёт ни один кортеж, т.к. ни для одного из кортежей, соответствующих схеме отношения-результата (Номер зачётной книжки, Фамилия, Оценка) не будет найдено соответствия всем предметам из отношения S .

В качестве примера рассмотрим один из таких кортежей: 123456-Савицкий-3. Чтобы этот кортеж мог попасть в результат, его декартово произведение с отношением-делителем, т.е. сочетание с каждым предметом, должно содержаться в отношении-делимом. Но в отношении «Вedomость» есть сочетание только с физикой. Кортежи 123456-Савицкий-Математика-3 и 123456-Савицкий-Русский язык-3 отсутствуют. Аналогичные рассуждения можно привести для всех остальных кортежей.

Поэтому перед выполнением операции деления нам необходимо с помощью взятия проекции избавиться от мешающего атрибута «Оценка» (которую по условию задачи и не требовалось включать в результат).

Разделив результат проекции отношения «Вedomость» по атрибутам «Номер зачётной книжки», «Фамилия» и «Предмет» на отношение «Предмет» мы получим требуемый список студентов, успешно сдавших все экзамены (рис. 4.19).

Отношение

$\Pi_{\text{Номер зачётной книжки, Фамилия, Предмет (Ведомость)}}$

Номер зачётной книжки	Фамилия	Предмет
123456	Савицкий	Физика
123457	Иванова	Математика
123458	Паулайнен	Математика
123456	Савицкий	Математика
123458	Паулайнен	Русский язык
123456	Савицкий	Русский язык

Отношение Предмет

Предмет
Физика
Математика
Русский язык

Отношение $\Pi_{\text{Номер зачётной книжки, Фамилия, Предмет (Ведомость)}} \div \text{Предмет}$

Номер зачётной книжки	Фамилия
123456	Савицкий

Рис. 4.19. Пример деления отношений

Таким образом, все операции реляционной алгебры можно условно разделить на две категории: *зависимые* реляционные операторы (выражаются через другие реляционные операторы) и *примитивные* (нельзя выразить через другие реляционные операторы).

К зависимым относят:

Оператор соединения – определяется через операторы декартова произведения и выборки (для оператора естественного соединения добавляется оператор проекции).

Оператор пересечения – выражается через вычитание.

Оператор деления – выражается через операторы вычитания, декартового произведения и проекции.

К примитивным относят: объединение, вычитание, декартово произведение, выборка и проекция.