



УНИВЕРСИТЕТ
СИНЕРГИЯ

КАФЕДРА ИНФОРМАЦИОННЫХ СИСТЕМ

УПРАВЛЕНИЕ ДАННЫМИ

Программа подготовки бакалавров по направлению
«Информационные системы и технологии»

Глушков Сергей Владимирович

Доцент, к.в.н., доцент



Реляционные БД

Теоретической основой этой модели стала теория отношений, основу которой заложили два логика — американец Чарльз Содерс Пирс (1839-1914) и немец Эрнст Шредер (1841-1902).

В руководствах по теории отношений было показано, что множество отношений замкнуто относительно некоторых специальных операций, то есть образует вместе с этими операциями абстрактную алгебру. Это важнейшее свойство отношений было использовано в реляционной модели для разработки языка манипулирования данными, связанного с исходной алгеброй.

- Американский математик Э. Ф. Кодд в 1970 году впервые сформулировал основные понятия и ограничения реляционной модели, ограничив набор операций в ней семью основными и одной дополнительной операцией.

Основной структурой данных в модели является отношение, именно поэтому модель получила название *реляционной* (от английского *relation* — отношение).

- *N*-арным отношением R называют подмножество декартова произведения $D_1 \times D_2 \times \dots \times D_n$ множеств D_1, D_2, \dots, D_n ($n > 1$), необязательно различных. Исходные множества D_1, D_2, \dots, D_n называют в модели *доменами*.
- $\{ R \subseteq D_1 \times D_2 \times \dots \times D_n \}$
- где $D_1 \times D_2 \times \dots \times D_n$ — полное декартово произведение

Полное декартово произведение — это набор всевозможных сочетаний из n элементов каждое, где каждый элемент берется из своего домена.

D1 содержит три фамилии, D2 — набор из двух учебных дисциплин и D3 — набор из трех оценок

- $D1 = \{\text{Иванов, Крылов, Степанов}\};$
- $D2 = \{\text{Теория автоматов, Базы данных}\};$
- $D3 = \{3, 4, 5\}$
- Тогда полное декартово произведение содержит набор из 18 троек, где первый элемент — это одна из фамилий, второй — это название одной из учебных дисциплин, а третий — одна из оценок.

- Отношение R моделирует реальную ситуацию и оно может содержать, допустим, только 5 строк, которые соответствуют результатам сессии (Крылов экзамен по "Бадам данных" еще не сдавал):
 - <Иванов, Теория автоматов, 4>;
 - <Крылов, Теория автоматов, 5>;
 - <Степанов, Теория автоматов, 5>;
 - <Иванов, Базы данных, 3>;
 - <Степанов, Базы данных, 4>;

R		
Фамилия	Дисциплина	Оценка
Иванов	Теория автоматов	4
Иванов	Базы данных	3
Крылов	Теория автоматов	5
Степанов	Теория автоматов	5
Степанов	Базы данных	4

Данная таблица обладает рядом специфических свойств:

- В таблице нет двух одинаковых строк.
- Таблица имеет столбцы, соответствующие атрибутам отношения.
- Каждый атрибут в отношении имеет уникальное имя.
- Порядок строк в таблице произвольный.

- Вхождение домена в отношение принято называть *атрибутом*.
- Строки отношения называются *кортежами*.
- Количество атрибутов в отношении называется степенью, или рангом, отношения.

Следует заметить, что в отношении не может быть одинаковых кортежей, это следует из математической модели: отношение — это подмножество декартова произведения, а в декартовом произведении все n -ки различны.

- В соответствии со свойствами отношений два отношения, отличающиеся только порядком строк или порядком столбцов, будут интерпретироваться в рамках реляционной модели как одинаковые

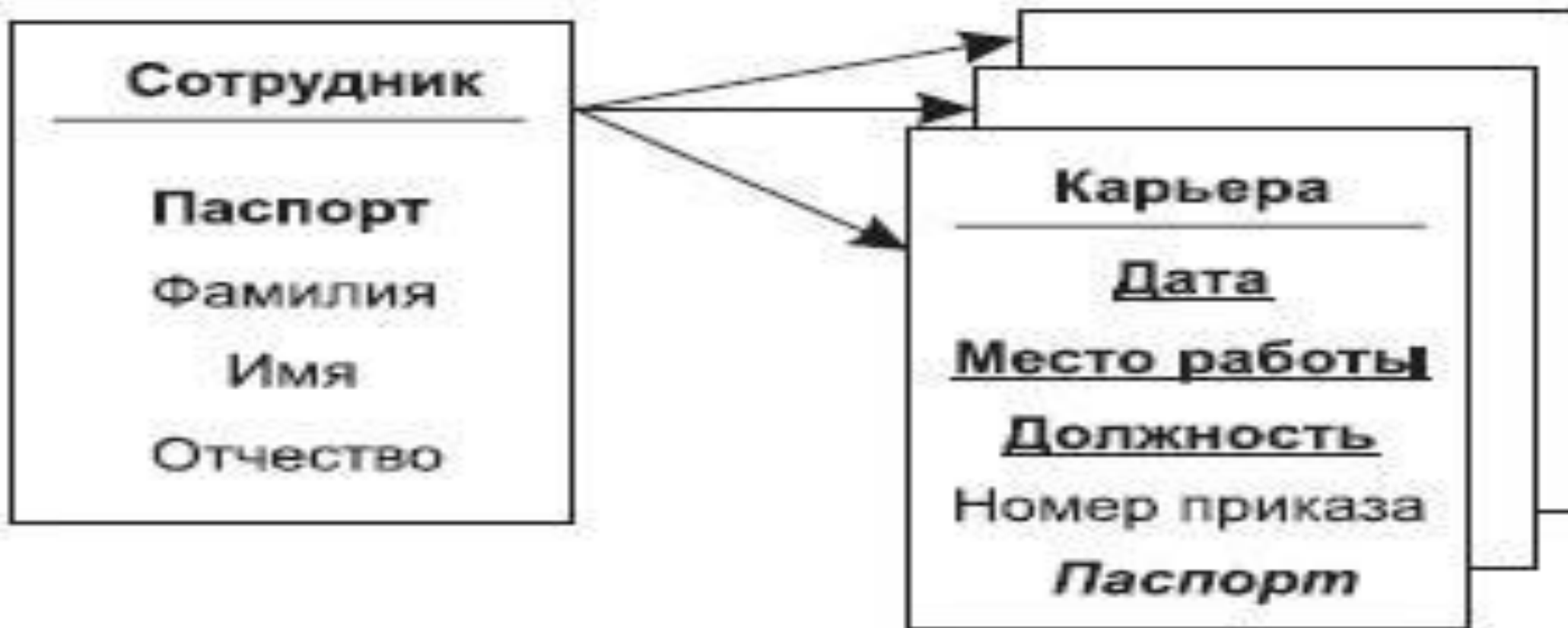
R1		
Дисциплина	Фамилия	Оценка
Теория автоматов	Крылов	5
Теорияавтоматов	Степанов	5
Теория автоматов	Иванов	4
Базыданных	Иванов	3
Базы данных	Степанов	4

Схемой отношения R

- называется перечень имен атрибутов данного отношения с указанием домена, к которому они относятся:
- $\{ S_{\{R\}} = (A_{\{1\}}, A_{\{2\}}, \dots, A_{\{n\}}), A_{\{i\}} \subseteq D_{\{i\}} \}$.
- Если атрибуты принимают значения из одного и того же домена, то они называются $\{ \theta \}$ - *сравнимыми*, где $\{ \theta \}$ — множество допустимых операций сравнения, заданных для данного домена.

- Схемы двух отношений называются *эквивалентными*, если они имеют одинаковую степень и возможно такое упорядочение имен атрибутов в схемах, что на одинаковых местах будут находиться сравнимые атрибуты, то есть атрибуты, принимающие значения из одного домена.

Связь между основным и подчиненным отношениями



Операции над отношениями. Реляционная алгебра

алгеброй называется множество объектов с заданной на нем совокупностью операций, замкнутых относительно этого множества, называемого ***основным множеством***

- Основным множеством в реляционной алгебре является множество отношений .

8 операций

- теоретико-множественные операции(4)
- *Три первые теоретико-множественные операции являются бинарными, то есть в них участвуют два отношения и они требуют эквивалентных схем исходных отношений.*
- специальные операции

- *Объединением* двух отношений называется отношение, содержащее множество кортежей, принадлежащих либо первому, либо второму исходным отношениям, либо обоим отношениям одновременно
- Пусть заданы два отношения $R_1 = \{ r_1 \}$, $R_2 = \{ r_2 \}$, где r_1 и r_2 - соответственно кортежи отношений
- $\{ R_1 \cup R_2 = \{ r \mid r \in R_1 \vee r \in R_2 \} \}$.
- Здесь r — кортеж нового отношения, $\{ \vee \}$ — операция логического сложения "ИЛИ".

- *Пересечением* отношений называется отношение, которое содержит множество кортежей, принадлежащих одновременно и первому и второму отношениям. R_1 и R_2 :
- $$R_3 = R_1 \cap R_2 = \{ r \mid r \in R_1 \wedge r \in R_2 \}$$
- *Разностью* отношений R_1 и R_2 называется отношение, содержащее множество кортежей, принадлежащих R_1 и не принадлежащих R_2 :
- $$R_5 = R_1 \setminus R_2 = \{ r \mid r \in R_1 \wedge r \notin R_2 \}$$

- $R_1 = (\text{ФИО}, \text{Паспорт}, \text{Школа}) ;$
- $R_2 = (\text{ФИО}, \text{Паспорт}, \text{Школа}) ;$
- $R_3 = (\text{ФИО}, \text{Паспорт}, \text{Школа}).$

- 1. Список абитуриентов, которые поступали два раза и не поступили в вуз.
- $\setminus [R = R_{\{1\}} \setminus \cap R_{\{2\}} \setminus \text{setminus} R_{\{3\}} \setminus]$
- 2. Список абитуриентов, которые поступили в вуз с первого раза, то есть они сдавали экзамены только один раз и сдали их так хорошо, что сразу были зачислены в вуз.
- $\setminus [R = (R_{\{1\}} \setminus \text{setminus} R_{\{2\}} \setminus \cap R_{\{3\}}) \setminus \cup (R_{\{2\}} \setminus \text{setminus} R_{\{1\}} \setminus \cap R_{\{3\}}) \setminus]$
- 3. Список абитуриентов, которые поступили в вуз только со второго раза.
- Прежде всего это те абитуриенты, которые присутствуют в отношениях R_1 и R_2 , потому что они поступали два раза, и присутствуют в отношении R_3 , потому что они поступили.
- $\setminus [R = R_{\{1\}} \setminus \cap R_{\{2\}} \setminus \cap R_{\{3\}} \setminus]$
- 4. Список абитуриентов, которые поступали только один раз и не поступили.
- Это прежде всего те абитуриенты, которые присутствуют в R_1 и не присутствуют в R_2 , и те, кто присутствуют в R_2 и не присутствуют в R_1 . И разумеется, никто из них не присутствует в R_3 .
- $\setminus [R = (R_{\{1\}} \setminus \text{setminus} R_{\{2\}}) \setminus \cup (R_{\{2\}} \setminus \text{setminus} R_{\{1\}}) \setminus \text{setminus} R_{\{3\}} \setminus]$

- **Операции объединения, пересечения и разности применимы только к отношениям с эквивалентными схемами**

- *Сцеплением, или конкатенацией*, кортежей $s = \langle s_1, s_2, \dots, s_n \rangle$ и $q = \langle q_1, q_2, \dots, q_m \rangle$ называется кортеж, полученный добавлением значений второго в конец первого. Сцепление кортежей s и q обозначается как (s, q) .
- $(s, q) = \langle s_1, s_2, \dots, s_n, q_1, q_2, \dots, q_m \rangle$
- Здесь n — число элементов в первом кортеже s , m — число элементов во втором кортеже q .

- *Расширенным декартовым произведением* отношения R_1 степени n со схемой
- $S_{R1} = (A_1, A_2, \dots, A_n),$
- и отношения R_2 степени m со схемой
- $S_{R2} = (B_1, B_2, \dots, B_m),$
- называется отношение R_3 степени $n+m$ со схемой
- $S_{R3} = (A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_m),$

Специальные операции реляционной алгебры

- Первой специальной операцией реляционной алгебры является *горизонтальный выбор*, или операция *фильтрации*, или операция *ограничения отношений*.

- Пусть a — булевское выражение, составленное из термов сравнения с помощью связок И (\wedge), ИЛИ (\vee), НЕ (\neg) и, возможно, скобок. В качестве термов сравнения допускаются:
 - **терм A θ a** ,
 - где A — имя некоторого атрибута, принимающего значения из домена D ; a — константа, взятая из того же домена D , $a \in D$; θ — одна из допустимых для данного домена D операций сравнения;
 - **терм A θ B** ,
 - где A, B — имена некоторых θ -сравнимых атрибутов, то есть атрибутов, принимающих значения из одного и того же домена D .

- Тогда результатом операции выбора, или фильтрации, заданной на отношении R в виде булевского выражения, определенного на атрибутах отношения R , называется отношение
 - $\sigma_{\alpha}(R)$,
- включающее те кортежи из исходного отношения, для которых истинно условие выбора или фильтрации:
- $\sigma_{\alpha}(R) = \{ r \mid r \in R \wedge \alpha(r) = \text{"Истина"} \}$

операция проектирования

- Пусть R — отношение, $S_R = (A_1, \dots, A_n)$ — схема отношения R .
- Обозначим через B подмножество $[A_i]$; $[B \subseteqq [A_i]]$.
- При этом пусть B^1 — множество атрибутов из $\{A_i\}$, не вошедших в B .
- Если $B = \{A_i^1, A_i^2, \dots, A_i^k\}$, $B = \{A_j^1, A_j^2, \dots, A_j^k\}$ и $[r = \langle a^{\{1\}}_i, a^{\{2\}}_i, \dots, a^{\{k\}}_i \rangle, a^{\{k\}}_i \in A^{\{k\}}_{ii}]$,
- то $r [B]$, $s = \langle a^1_j, a^2_j, \dots, a^m_j \rangle$; $[a^{\{m\}}_j \in A^{\{m\}}_j]$.

- Проекцией отношения R на набор атрибутов V , обозначаемой $R[V]$, называется отношение со схемой, соответствующей набору атрибутов V $S_{R[V]} = V$, содержащему кортежи, получаемые из кортежей исходного отношения R путем удаления из них значений, не принадлежащих атрибутам из набора V .

- $R[V] = \{ r[V] \}$

- Операция проектирования, называемая иногда также операцией вертикального выбора, позволяет получить только требуемые характеристики моделируемого объекта. Чаще всего операция проектирования употребляется как промежуточный шаг в операциях горизонтального выбора, или фильтрации. Кроме того, она используется самостоятельно на заключительном этапе получения ответа на запрос.

операция *условного соединения*.

- В отличие от рассмотренных специальных операций реляционной алгебры: фильтрации и проектирования, которые являются унарными, то есть производятся над одним отношением, операция *условного соединения* является бинарной, то есть исходными для нее являются два отношения, а результатом — одно.

операция деления.

- даны два отношения R и T соответственно со схемами:
- $S_R = (A_1, A_2, \dots, A_k); S_T = (B_1, B_2, \dots, B_m);$
- A и B - наборы атрибутов этих отношений, одинаковой длины (без повторений);
- операция деления ставит в соответствие отношениям R и T отношение $Q = R[A:B]T$, кортежи которого являются теми элементами проекции $R[A^1]$, для которых $T[B]$ входит в построенные для них множество образов:
- $$\{ R[A:B]T = \{ r \mid r \in R[A^1] \wedge T[B] \subseteq \{ y \mid y \in R[A] \wedge (r, y) \in R \} \} \}$$

пример

- $R_1 = \langle \text{ФИО, Дисциплина, Оценка} \rangle$;
- $R_2 = \langle \text{ФИО, Группа} \rangle$;
- $R_3 = \langle \text{Группы, Дисциплина} \rangle$,
- где R_1 — информация о попытках (как успешных, так и неуспешных) сдачи экзаменов студентами;
- R_2 — состав групп;
- R_3 — список дисциплин, которые надо сдавать каждой группе.

- Список студентов, которые сдали экзамен по БД на "отлично". Результат может быть получен применением операции фильтрации по сложному условию к отношению R_1 и последующим проектированием на атрибут "ФИО" (нам ведь требуется только список фамилий).
- $\pi [S = (R_{1}[Оценка = 5 \wedge Дисциплина = "БД"])[ФИО]] ;$

- Список тех, кто должен был сдавать экзамен по БД, но пока еще не сдавал. Сначала найдем всех, кто должен был сдавать экзамен по БД. В отношении R_3 находится список всех дисциплин, по которым каждая группа должна была сдавать экзамены, ограничим перечень дисциплин только "БД". Для того чтобы получить список студентов, нам надо соединить отношение R_3 с отношением R_2 , в котором определен список студентов каждой группы.
- $\left[R_4 = (R_2 \bowtie R_3) \text{.НомерГруппы} = R_2 \text{.НомерГруппы} \wedge R_3 \text{.Дисциплина} = \text{"БД"} \right] R_3 \text{[ФИО]} \setminus ;$

- Теперь получим список всех, кто сдавал экзамен по "БД" (нас пока не интересует результат сдачи, а интересует сам факт попытки сдачи, то есть присутствие в отношении R_1):
- $R_5 = (R_1 [\text{Дисциплина} = \text{"БД"}])[\text{ФИО}]$;
- и, наконец, результат — все, кто есть в первом множестве, но не во втором:
- $S = R_4 \setminus R_5$;

- Список несчастных, имеющих несколько двоек:
- $\setminus [S = (R_{\{1\}}[R_{\{1\}}. \text{ФИО} = R'_{\{1\}}. \text{ФИО} \setminus \wedge R_{\{1\}}.$
 Дисциплина $\setminus \neq R'_{\{1\}}.$ Дисциплина $\setminus \wedge R_{\{1\}}.$ Оценка
 $< 2 \setminus \wedge R'_{\{1\}}.$ Оценка $< 2] R'_{\{1\}}) [\text{ФИО}] \setminus$
- Список круглых отличников. Строим список всех пар <студент—дисциплина>, которые в принципе должны быть сданы:
- $R_4 = (R_2 [R_2 \text{ Группа} = R_3. \text{Группа}] R_3) [\text{ФИО}, \text{Дисциплина}] ;$

- Строим список пар <студент—дисциплина>, где получена оценка "отлично":
- $R_5 = (R_1[\text{Оценка} = 5])[\text{ФИО}, \text{Дисциплина}]$;
- Строим список студентов, что-либо не сдавших на "отлично":
- $R_6 = (R_4 \setminus R_5)[\text{ФИО}]$.
- $R_2[\text{ФИО}] \setminus R_6$

