КАФЕДРА ИНФОРМАЦИОННЫХ СИСТЕМ

УПРАВЛЕНИЕ ДАННЫМИ

Программа подготовки бакалавров по направлению «Информационные системы и технологии»

Глушков Сергей Владимирович

Доцент, к.в.н., доцент


Модели данных
Инфологические модели
Даталогические модели
Физические модели
Диаграммы Бахмана
Модель сущность связь (ER)
Докумен-тальные модели
Фактогра-фические модели
Основанные на файловых структурах
Основанные на странично-сегментной организации
Ориентиро-ванные на формат документа
Дескрип-торные модели
Тезаурусные модели
Теоретико-графовые
Теоретико-множественные
Объектно-ориентиро-ванные
Иерархическая
Сетевая
Реляцион-ная
Бинарных ассоциаций

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 D1x D2x … xDn множеств D1, D2, ..., Dn (n > 1), необязательно различных. Исходные множества D1, D2, ..., Dn называют в модели *доменами*.
- $R \subseteq D_{1} \times D_{2} \times \dots \times D_{n}$
- где D1 x D2 x … xDn — 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 = {Иванов, Крылов, Степанов};
- D2 = {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 "Базам 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

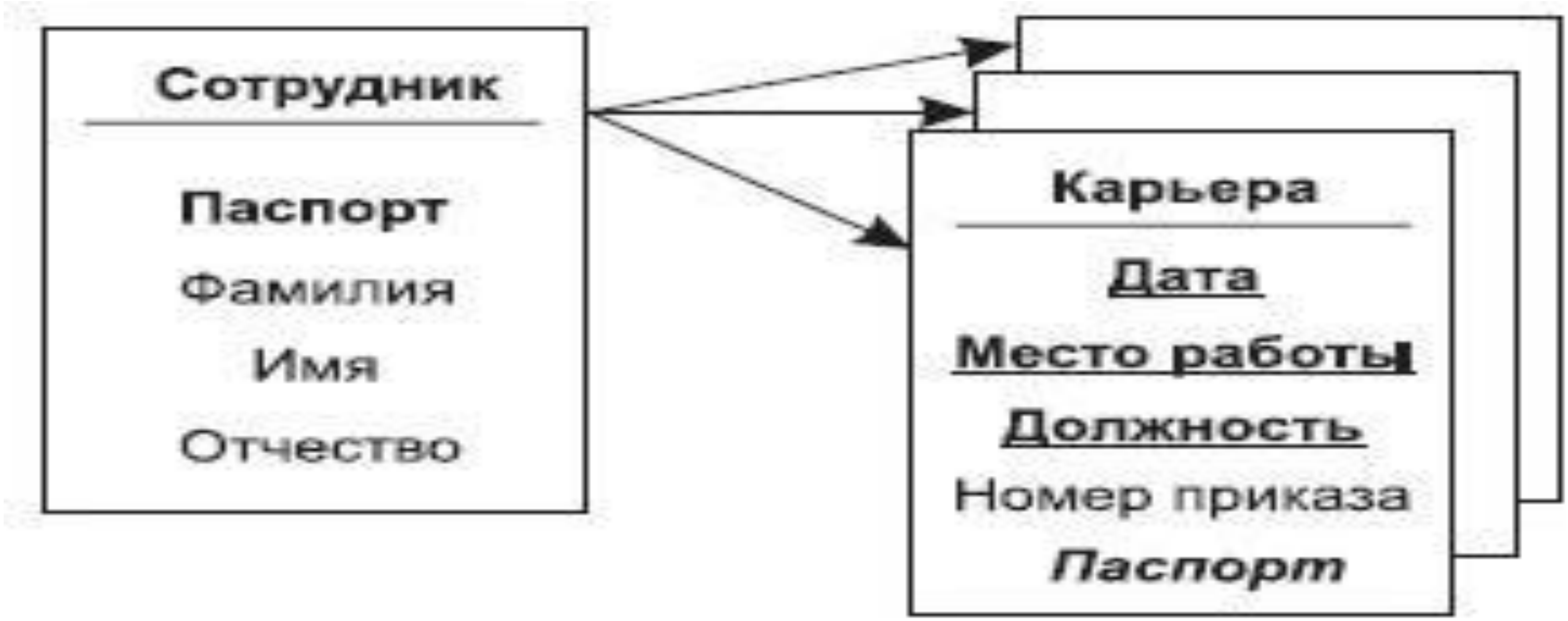
- называется перечень имен атрибутов данного отношения с указанием домена, к которому они относятся:

- $$S_{R} = (A_{1}, A_{2}, A_{n}), A_{i} \subseteq D_{i}$$.

- Если атрибуты принимают значения из одного и того же домена, то они называются θ - *сравнимыми*,где θ — 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 .

УНИВЕРСИТЕТ СИНЕРГИЯ

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 R1 = { r1 } , R2 = { r2}, где r1 и r2 - соответственно кортежи отношений
- $$ R_{1} \cup R_{2} = \{ r | r \in R_{1} \vee r \in R_{2}\} $$.
- 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. R1 и R2:
- $$ R_{3} = R_{1} \cap R2 =\{ r \mid r \in R1 \wedge r \in R_{2} \} $$

- *Разностью* отношений R1 и R2 называется отношение, содержащее множество кортежей, принадлежащих R1 и не принадлежащих R2:
- $$ R_{5} =R_{1} \setminus R_{2} =\{ r \mid r \in R1 \wedge r \notin R_{2}\} $$

- R_1= (ФИО, Паспорт, Школа) ;
- R_2= (ФИО, Паспорт, Школа) ;
- R_3= (ФИО, Паспорт, Школа).

- 1. Список абитуриентов, которые поступали два раза и не поступили в вуз.
- $R = R_{1} \cap R_{2} \setminus R_{3}$
- 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.
- $R = (R_{1} \setminus R_{2} \cap R_{3}) \cup (R_{2} \setminus R_{1} \cap R_{3})$
- 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.
- $R=R_{1} \cap R_{2} \cap R_{3}$
- 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.
- $R = (R_{1} \setminus R_{2}) \cup (R_{2} \setminus R_{1}) \setminus R_{3}$

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

- *Сцеплением*,или *конкатенацией*,кортежей $c = \langle c_1, c_2, \ldots, c_n \rangle$ и $q = \langle q_1, q_2, \ldots, q_m \rangle$ называется кортеж, полученный добавлением значений второго в конец первого. Сцепление кортежей c и q обозначается как (c , q).
- $(c, q) = \langle c_1, c2, \ldots , c_n, q_1, q_2, \ldots, q_m \rangle$

- Здесь n — число элементов в первом кортеже c, m — число элементов во втором кортеже q.

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

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 oc a,
- где A — имя некоторого атрибута, принимающего значения из домена D ; a — константа, взятая из того же домена D, \[a \in D \] ; oc — одна из допустимых для данного домена D операций сравнения;
- терм A oc B,
- где A, B — имена некоторых \[\theta \] -сравнимых атрибутов, то есть атрибутов, принимающих значения из одного и то же домена D.

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

операция проектирования

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

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

- R[B] = { r[B] }

- 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 = <ФИО, Дисциплина, Оценка> ;
- R_2 = <ФИО, Группа> ;
- R_3 = < Группы, Дисциплина>,
- *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).
- \[S = (R_{1}[Оценка = 5 \wedge Дисциплина = "БД"])[ФИО] \] ;

- 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.
- $$R_{4} = (R_{2}[R_{3}.\text{НомерГруппы} = R_{2}.\text{НомерГруппы} \wedge R_{3}.\text{Дисциплина} = \text{"БД"}] R_{3})[\text{ФИО}]$$;

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

- Строим список пар <студент—дисциплина>, где получена оценка "отлично":
- $R5 = (R_1[Оценка = 5])[ФИО, Дисциплина]$;
- Строим список студентов, что-либо не сдавших на "отлично":
- $R_6 = (R_4 \setminus R5)[ФИО]$.
- $R_2[ФИО] \setminus R_6$