

Реляционная алгебра (окончание)

Бессарабов Н.В.

bes@fpm.kubsu.ru

2017 г.

Реляционная модель данных

Уже говорилось о том, что любая модель данных состоит из трех частей:

- Структурной;
- Целостной;
- Манипуляционной.

Особенности реляционной модели:

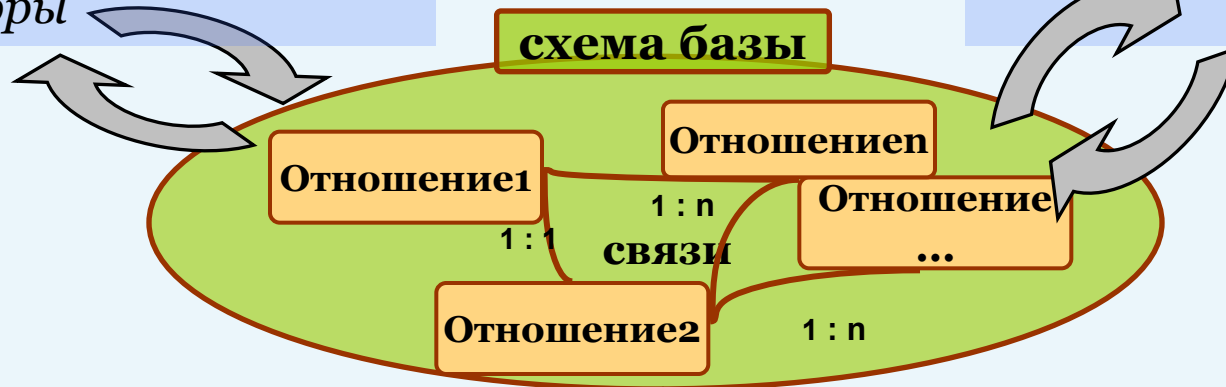
- Схема базы образуется единственным источником данных – отношениями -- и ограниченным набором связей между отношениями имеющими тип “один-к-одному” и “один-ко-многим”;
- Отношения строятся только на скалярных predetermined типах данных;
- Используется два теоретически эквивалентных способа манипулирования данными – **реляционная алгебра** и **реляционные исчисления**.

Замечание: В реляционной модели под манипулированием данными понимается построение новых временных отношений из набора уже имеющихся. Средств для создания отношений, не выводимых из имеющихся, и для изменения состояния отношений (т.е. заполнения их кортежами или изменения кортежей) не существует.

Особенности реляционной модели

Манипуляции данными с помощью *реляционной алгебры*

Манипуляции данными с помощью *реляционных исчислений*



Используют только простые типы данных

В реляционной модели это единственный источник данных

Манипулирование данными в реляционной модели это построение новых отношений из уже имеющих. Отношения не выводимые из имеющих создать нельзя. Нет заполнения отношений кортежами.

Реляционная алгебра

Определяется на конечном множестве отношений с фиксированной **сигнатурой** и **конечным** числом кортежей. Поскольку сигнатуры отношений могут не совпадать, реляционная алгебра **многосортна**, сами отношения и кортежи разных отношений могут быть не сравнимы.

Отношение r определяется своей схемой R . Набор записей в отношении определяет его состояние. При этом повторяющиеся кортежи отсутствуют.

Замечание: Ещё раз обратим внимание на то, что набор схем отношений предполагается заданным заранее. Реляционная алгебра не изменяет его и не может изменять состояние отношений, то есть вводить, удалять и изменять записи. Манипуляции данными создают временные, не сохраняемые отношения.

Операции реляционной алгебры

Перечень операций:

- Проекция
- Естественное соединение
- \bowtie - соединение
- Декартово произведение
- Селекция
- Булевы операции
- Частное
- Переименование атрибутов

Переименование атрибутов –
самая необычная операция !

Две операции уже рассмотрены в предыдущей лекции:

- 1) Проекция обозначаемая $proj_x(r)$.
- 2) Естественное соединение обозначаемое $join(r_1, r_2)$, $join_{=A}(r_1, r_2)$ или $r_1 join r_2$.

⋈ - соединение

Определение: Пусть даны отношения r_1, r_2 со схемами

$R_1(A_1, \dots, A_k, B_1, \dots, B_l),$

$R_2(C_1, \dots, C_m, D_1, \dots, D_n),$ соответственно;

θ -- оператор сравнения на группах атрибутов A и C .

Тогда θ - **соединение** отношений r_1 и r_2 есть отношение r_3 со схемой

$R_3(A_1, \dots, A_k, B_1, \dots, B_l, C_1, \dots, C_m, D_1, \dots, D_n),$

полученной объединением атрибутов схем R_1 и R_2 **без повторения**. Записи r_3 получаются конкатенацией тех записей из r_1 и r_2 , у которых значения группы столбцов A в r_1 и группы столбцов C в r_2 находятся в отношении θ (удовлетворяют θ).

Обозначение: $\text{join}_{A\theta C}(r_1, r_2)$

Замечание: Очевидно, если θ есть равенство “=” и $A \equiv C$ получим естественное соединение со схемой

$R_3(A_1, \dots, A_k, B_1, \dots, B_l, D_1, \dots, D_n).$

Пример ∩ - соединения

Исходные отношения

employee

ename	job	sal
-------	-----	-----

salgrade

grade	losal	hisal
-------	-------	-------

Условие $\cap = losal \leq sal \leq hisal$

Пример заполнения исходных отношений

employee

ename	job	sal
Иванов	прогр.	4500
Петров	нач. отд.	8200

salgrade

grade	losal	hisal
1	3001	5000
2	5001	8000
3	8001	15000

Результат \cap -
соединения

ename	job	sal	grade	losal	hisal
Иванов	прогр.	4500	1	3001	5000
Петров	нач. отд.	8200	3	8001	15000

Декартово произведение

Определение: Декартовым произведением отношений r и s арностей k_r и k_s , с непересекающимися множествами атрибутов, соответственно, R и S , называется отношение $t = r \times s$ арности $k_r + k_s$, состоящее из кортежей, первые k_r компонентов которых есть кортежи из r , а последние k_s компонентов выбираются из s . Иначе говоря, кортежи t образованы конкатенацией каждого кортежа из r с каждым кортежем из s . Поэтому, если в текущем состоянии r и s имеют n_r и n_s кортежей, то в t их $n_r \times n_s$.

Замечание: В одном отношении недопустим повтор имен. Поэтому, в частности, не существует декартов квадрат. При соединении отношений с одноименными атрибутами некоторые из них могут быть переименованы исходя из семантики данных и соединения.

Пример декартова произведения

r:

A	B	C
a ₁	b ₁	c ₁
a ₂	b ₁	c ₂
a ₃	b ₃	c ₂

s:

D	E
d ₁	e ₁
d ₂	e ₂

$r \times s$:

A	B	C	D	E
a ₁	b ₁	c ₁	d ₁	e ₁
a ₁	b ₁	c ₁	d ₂	e ₂
a ₂	b ₁	c ₂	d ₁	e ₁
a ₂	b ₁	c ₂	d ₂	e ₂
a ₃	b ₃	c ₂	d ₁	e ₁
a ₃	b ₃	c ₂	d ₂	e ₂

Селекция (выбор)

Определение: Пусть F – формула, образованная:

- операндами в виде констант или имен столбцов (номеров столбцов);
- операторами сравнения $<, =, >, \leq, \geq, \neq$;
- логическими операторами \wedge, \vee, \neg .

А если допустить значение null?

Тогда результат селекции $\text{sel}_F(r)$ есть множество кортежей t из r , для которых формула F истинна.

Пример:

r:

A	B	C
a_1	b_1	c_1
a_2	b_1	c_2
a_3	b_3	c_2

$\text{sel}_{C=c_2 \wedge B=b_1}(r)$:

A	B	C
a_2	b_1	c_2

Булевы операции

Два отношения r_1 и r_2 с одной и той же схемой R могут рассматриваться как подмножества множества всех возможных кортежей в схеме R . Поэтому к ним применимы булевы операции \cap , \cup , $-$

r_1 :

A	B	C
a_1	b_1	c_1
a_2	b_1	c_2
a_3	b_2	c_1

r_2 :

A	B	C
a_1	b_1	c_1
a_2	b_1	c_1
a_3	b_2	c_2

Замечание: Булевы операции могут применяться к совместимым отношениям, у которых атрибуты попарно имеют совместимые типы и общую семантику.

$r_1 \cap r_2$:

A	B	C
a_1	b_1	c_1

$r_1 \cup r_2$:

A	B	C
a_1	b_1	c_1
a_2	b_1	c_2
a_3	b_2	c_1
a_2	b_1	c_1
a_3	b_2	c_2

$r_1 - r_2$:

A	B	C
a_2	b_1	c_2
a_3	b_2	c_1

Дополнение

В определении дополнения возникают трудности. Пусть $\text{dom}(R)$ есть множество всех возможных кортежей над атрибутами схемы R с определенным для каждого атрибута доменом.

Если хотя бы один домен бесконечен, то полное отношение r^* , включающее все элементы из $\text{dom}(R)$ не будет отношением в понимании реляционной алгебры.

Не будет отношением и дополнение к конечному отношению r :

$$\overline{\exists} r^* - r$$

Частное

Определение: Пусть даны:

- отношение r с арностью k_r и схемой R
- и
- отношение s с арностью $k_s < k_r$ и схемой S , которая не пуста, то есть $S \neq \emptyset$, и является собственным подмножеством схемы R , то есть $S \subset R$.

Тогда **частным** называется отношение $r \div s$ арности $k_r - k_s$, которое:

- содержит столбцы отношения r отсутствующие в s ;
- часть записи r включается в $r \div s$ если в r она сцеплена с каждой записью из s .

Замечание: Смысл введения этой операции станет понятен позднее при изучении многозначных функциональных зависимостей (MV-зависимостей).

Пример частного

r:

A	B	C	D
a ₁	b ₂	c ₃	d ₁
a ₁	b ₂	c ₄	d ₃
a ₂	b ₃	c ₄	d ₃
a ₃	b ₄	c ₃	d ₁
a ₃	b ₄	c ₄	d ₃

s:

C	D
c ₃	d ₁
c ₄	d ₃

r ÷ s:

A	B
a ₁	b ₂
a ₃	b ₄

Обозначение: r division s или
division(r,s)

или $r \div s$

Совместимость отношений и переименование атрибутов

Теоретико-множественные операторы объединение, пересечение и разность требуют, чтобы отношения – операнды были совместимы, то есть относились к элементам одного сорта. Это означает, что отношения отличаются только именами и состояниями. Сигнатуры у них одинаковы, то есть количества атрибутов совпадают и атрибуты попарно совпадают по типам, а в простейшем случае, по именам.

Если же имена отношений и/или атрибутов не совпадают, необходимо установить соответствие между именами отношений или изменить некоторые из имён атрибутов.

В операциях соединений и декартовом произведении может появиться повтор одинаковых атрибутов, что делает невозможным выполнение операции. И здесь переименование может позволить выполнение операции.

Итак, некоторые несовместимые отношения могут стать совместимыми после переименования атрибутов. Поэтому необходимо ввести операцию **переименования атрибутов**.

Замечание: Может быть не понятно, зачем нужны декартовы произведения. Необходимые разъяснения будут приведены на следующем слайде, слайде 18 и при изучении языка SQL.

Примеры переименования

Пример 1: Необходимо объединить отношения “Employee” и “Работники” для расчета суммарной заработной платы:

Employee(empno, ename, salary, mgr)

Работник(Тно, ФИО, зарплата, Тно_нач)

где Тно – табельный номер, Тно_нач -- табельный номер начальника.

Выполняем переименования атрибутов: Тно → empno, ФИО → ename, зарплата → salary, Тно_нач → mgr (типы и смысл соответствующих атрибутов считаются одинаковыми).

Операция переименования атрибутов может выглядеть так:

```
[имя_отношения] RENAME список_старых_атрибутов AS  
список_новых_атрибутов
```

Пример 2: Переименование атрибутов необходимо для объединения отношения с собой. Скажем, необходимо выбрать всех сотрудников и их непосредственных начальников. Ответ можно получить из декартова произведения отношения “Employee” с собой после переименования.

Его атрибуты

(empno, ename, salary, mgr, empno_mgr, ename_mgr, salary_mgr, mgr_mgr)

Независимые операции реляционной алгебры

Объединение, вычитание, декартово произведение, селекция и проекция независимые (примитивные) операции - их нельзя выразить друг через друга.

- Декартово произведение и соединения -- единственные операции, *увеличивающие количество атрибутов*. Поэтому они не выразимы через остальные операции, не обладающие этим свойством. Поскольку декартово произведение в некотором смысле проще, а соединение представляется как его подмножество, следует считать декартово произведение независимой операцией.
- Проекция - единственная операция, *уменьшающая количество атрибутов в одном операнде*. Поэтому её нельзя выразить через остальные операции, не обладающие этим свойством.
- Селекция - единственная операция, *сравнивающая атрибуты отношения*. Поэтому она не выразима через остальные операции, не обладающие этим свойством.
- Доказательство независимости объединения и вычитания не приводятся. Отметим только, что теоретико-множественные операции единственные требуют совместимости операндов, так что какие-то из них независимы.

Зависимые операции реляционной алгебры

Операции *соединения, пересечения и деления* можно выразить через другие реляционные операции

- Операция соединения определяется через операции декартового произведения и выборки.

$$\text{join}_F (r_1, r_2) = \text{sel}_F (r_1 \times r_2)$$

- Операция пересечения выражается через вычитание следующим образом:

$$r_1 \cap r_2 = r_1 - (r_1 - r_2)$$

- Оператор деления выражается через операторы вычитания, декартового произведения и проекции следующим образом:

$$r_1 \div r_2 = \text{proj}_X r_1 - \text{proj}_X ((\text{proj}_X r_1) \text{ join } r_2) - r_1$$

Реляционная алгебра. Перечень обозначений.

Обозначим:

- **U** – множество атрибутов (универсум),
- **D** – множество доменов,
- **dom** – полная функция $\text{dom} : \mathbf{U} \rightarrow \mathbf{D}$; назначает домен каждому атрибуту,
- **R** – множество всех возможных схем отношений на **U** ,
- $\mathcal{r} = \{r_1, \dots, r_p\}$ есть множество отношений r_i со схемами R_i , соответственно,
- θ - множество бинарных отношений, определенных на доменах из **D** содержащее, по крайней мере, отношение равенства и неравенства для каждого домена.

Реляционная алгебра. Определение.

Определение: Реляционной алгеброй над U , D , dom , R , r , θ называется семиместный кортеж

$$B = (U, D, dom, R, r, \theta, O),$$

где O – это множество операций, содержащее операции селекции, проекции, объединения, пересечения, разности, частного, естественного и \bowtie - соединения, а также операцию переименования атрибутов из U .

Примеры запросов (1/2)

Заданы отношения:

emp (empno, ename, job, sal, deptno)

dept (deptno, dname, loc)

Уточните семантику отношений!

В dept и emp имеется столбец deptno !!

Смысл имен с точки зрения предметной области:

emp от employee -- работник;

dept от department – отдел;

empno табельный номер;

ename имя работника (employee name);

job должность;

sal от salary --заработная плата;

deptno номер отдела;

dname название отдела;

loc местонахождение отдела.

Примеры запросов (2/2)

1. Выдать фамилии и должности лиц, получающих зарплату больше 1000:

```
proj {ename, job} (sel sal>1000 (emp))
```

2. Выдать список сотрудников в виде отношения с атрибутами: empno, ename, job, dname.

Первая неудачная попытка. Запрос с соединением:

```
proj {empno, ename, job, dname} (join deptno=deptno (emp, dept))
```

недопустим, так как в условии соединения deptno=deptno не понятно, из каких отношений берутся эти атрибуты. К тому же само условие получилось тождественно истинным.

Переименуем атрибуты таким образом: deptno из dept обозначим dept1, а deptno из emp оставим без изменения.

Правильный запрос:

```
proj {empno, ename, job, dname} (join deptno1=deptno (emp, dept))
```

Замечание: в реализациях можно использовать уточнение имени атрибута именем отношения. В нашем примере будет так: dept.deptno и emp.empno.

Сравнение отношений и их табличных реализаций в БД

Три отличия отношений от таблиц:

- В отношении нет одинаковых кортежей. Таблицы без ключа могут содержать одинаковые строки.
- Тело отношения есть множество и потому кортежи не упорядочены. Строки таблиц могут быть упорядочены. В этом случае одно отношение можно реализовать таблицами, отличающимися порядком строк.
- Атрибуты отношения определяются с уникальными в пределах отношения именами и потому не нуждаются в упорядочении. Столбцы таблиц могут быть упорядочены. В некоторых реализациях имена столбцов могут заменяться их номерами. Одно отношение можно реализовать таблицами, со столбцами записанными в разном порядке.

Отношения и таблицы. Термины.

Реляционный термин	"Табличный" термин
Схема реляционной базы данных	Схема базы данных
Отношение	Таблица
Семантика отношения	Семантика таблицы
Заголовок отношения	Заголовок таблицы
Тело отношения	Тело таблицы
Атрибут отношения	Столбец таблицы
Семантика атрибута	Семантика столбца
Кортеж отношения	Строка таблицы
Арность отношения	Количество столбцов таблицы
Типы данных и домены	Типы данных и домены

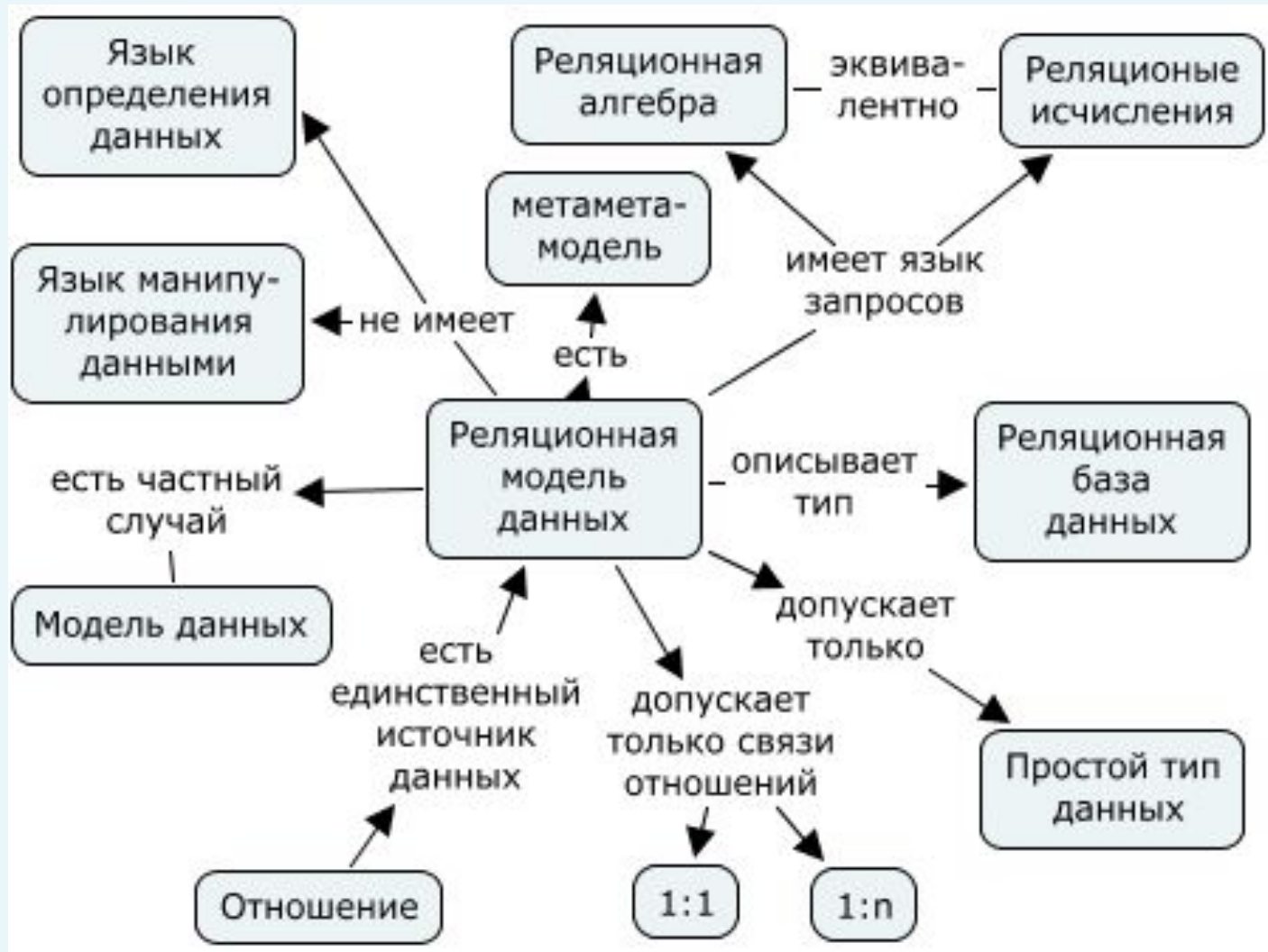
Заключение

- Выражения реляционной алгебры строятся на отношениях и возвращают отношения же. Отношения-результаты можно использовать как аргументы в других выражениях.
- Из-за необходимости использования декартова произведения при выполнении соединения таблицы представляющие промежуточные результаты могут иметь громадные размеры. Поэтому в реализациях СУБД реляционную алгебру в настоящее время не используют.
- Выделяются две группы операций: А зачем мы её изучаем?
 - Теоретико-множественные: объединение, пересечение, вычитание, декартово произведение.
 - Реляционные: выборка, проекция, селекция, соединение, частное.
- Для выполнения некоторых операций необходимо обеспечить совместимость отношений по сигнатуре. Для этого используют *переименование атрибутов и отношений*.
- Независимость операций. Операции *соединение, пересечение и частное* можно выразить через другие реляционные операции. Операции *объединение, вычитание, декартово произведение, селекция, проекция* нельзя выразить друг через друга.
- Реляционная алгебра это язык запросов. Выразить создание исходных отношений, заполнить их, изменить или удалить кортежи в этой алгебре нельзя.

Основные понятия (1/2)



Основные понятия (2/2)



Словарь студента (1/4)

- ❑ **Алгебра реляционная** – см. слайд 20
- ❑ **Дополнение** – теоретико-множественная операция, не используемая в реляционной алгебре
- ❑ **Запрос** – сообщение конечного пользователя или приложения, направляемое СУБД и активизирующее в системе базы данных действия, которые обеспечивают выборку, вставку, удаление или обновление указанных в нем данных. Для описания запросов используются языки запросов. (М.Р. Когаловский)
- ❑ **Исчисление реляционное** – специальная форма исчисления предикатов первого порядка, которая может использоваться как основа *декларативных языков запросов*. В таких языках запросы записываются в виде логической формулы, которая должна быть истинной для кортежей отношения, составляющих результат запроса. (М.Р. Когаловский)
- ❑ **Операции булевы** в реляционной алгебре определяются на наборах кортежей. Не всегда применимы из-за многосортности реляционной алгебры.
- ❑ **Операция переименования атрибутов:**
[нов_имя_отношения] RENAME список_старых_атрибутов AS список_новых_атрибутов

Словарь студента (2/4)

- ❑ **Операции зависимые** (в реляционной алгебре)- операции *соединения, пересечения и деления* - можно выразить через другие реляционные операции.
- ❑ **Операции независимые** - Объединение, вычитание, декартово произведение (увеличивает кол-во атрибутов), выборка (сравнивает атрибуты отношения) и проекция (уменьшает кол-во атрибутов) независимые (**примитивные операции**) - их нельзя выразить друг через друга.
- ❑ **Произведением декартовым** отношений r и s арностей k_r и k_s , с непересекающимися множествами атрибутов, соответственно, R и S , называется отношение $t = r \times s$ арности $k_r + k_s$, состоящее из кортежей, первые k_r компонентов которых есть кортежи из r , а последние k_s компонентов выбираются из s . Иначе говоря, кортежи t образованы конкатенацией каждого кортежа из r с каждым кортежем из s . Поэтому, если в текущем состоянии r и s имеют n_r и n_s кортежей, то в t их $n_r \times n_s$.
- ❑ **Проекция** - это набор унарных операций выбора подмножества столбцов отношений $\text{proj}_x(r)$, где R схема отношения r и $x \subseteq R$ – набор столбцов.

Словарь студента (3/4)

- ❑ **Селекция.** Пусть F – формула, образованная:
 - операндами в виде констант или имен столбцов (номеров столбцов)
 - операторами сравнения $<, =, >, \leq, \geq, \neq$
 - логическими операторами \wedge, \vee, \negТогда результат селекции $\text{sel}_F(r)$ есть множество кортежей t из r , для которых формула F истинна.
- ❑ **Сигнатура** отношения – число мест и список типов.
- ❑ **Совместимость операндов** - то есть принадлежность к элементам одного сорта. Совместимые отношения отличаются только именами и состояниями. Сигнатуры у них одинаковы.
- ❑ **Соединение естественное** – Пусть отношения r_1 и r_2 имеют схемы $R_1(A_1, \dots, A_k, B_1, \dots, B_n)$ и $R_2(A_1, \dots, A_k, C_1, \dots, C_m)$. Тогда естественное соединение (join) отношений r_1 и r_2 есть отношение r_3 со схемой $R_3(A_1, \dots, A_k, B_1, \dots, B_n, C_1, \dots, C_m)$ в котором каждая запись (экземпляр) получена конкатенацией каждой записи из r_1 с теми записями из r_2 , у которых совпадают значения в общих атрибутах A_1, \dots, A_k .

Словарь студента (4/4)

- **∅ - соединение.** Пусть даны отношения r_1, r_2 со схемами $R_1(A_1, \dots, A_k, B_1, \dots, B_l), R_2(C_1, \dots, C_m, D_1, \dots, D_n)$, соответственно;
 \emptyset -- оператор сравнения на группах атрибутов A и C .

Тогда \emptyset - **соединение** отношений r_1 и r_2 есть отношение r_3 со схемой $R_3(A_1, \dots, A_k, B_1, \dots, B_l, C_1, \dots, C_m, D_1, \dots, D_n)$, полученной объединением атрибутов схем R_1 и R_2 **без повторения**. Записи r_3 получаются конкатенацией тех записей из r_1 и r_2 , у которых значения группы столбцов A в r_1 и группы столбцов C в r_2 находятся в отношении \emptyset (удовлетворяют \emptyset).

$(\text{join}_{A \emptyset C}(r_1, r_2))$

- **Частное.** Пусть даны:

отношение r с арностью k_r и схемой R и отношение s с арностью $k_s < k_r$ и схемой S , причем $S \subset R$ и $S \neq \emptyset$. Тогда **частным** называется отношение $r \div s$ арности $k_r - k_s$, которое:

- содержит столбцы отношения r отсутствующие в s ;
- часть записи r включается в $r \div s$ если в r она сцеплена с каждой записью из s .