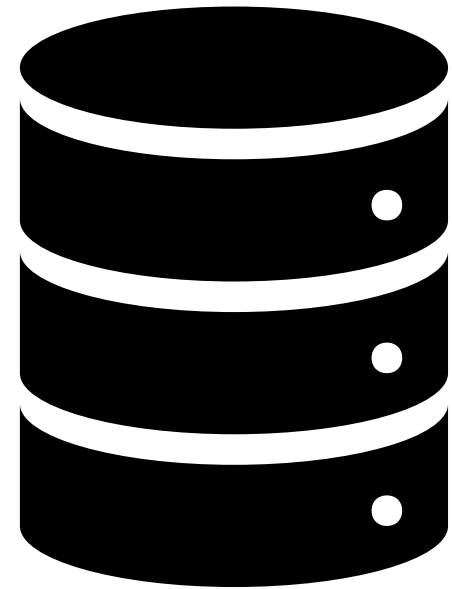


Базы данных

Лекция 1

Введение в базы данных



✉ bakanchevn@gmail.com
📧 @nikbeesti

ФПМИ МФТИ,
2022

СТРУКТУРА КУРСА

- Лекции / Семинары
- Семинары
 - Домашняя работа (интересная и не очень)
 - Контрольные работы (сложные и не очень)
 - Практические задачи (интеллектуальные и не очень)
- Лекции
 - Теория и иногда практика



О ЧЕМ КУРС?

- Предпосылки к созданию баз данных (БД)
- Реляционная алгебра и реляционные БД
- Язык запросов SQL
- Системы управления базами данных (СУБД)
- Троичная логика
- Моделирование БД
- Продвинутое запросы SQL
- Оптимизация запросов
- Основы хранилищ данных (DWH) и современное применение БД
- Нереляционные БД (NoSQL)



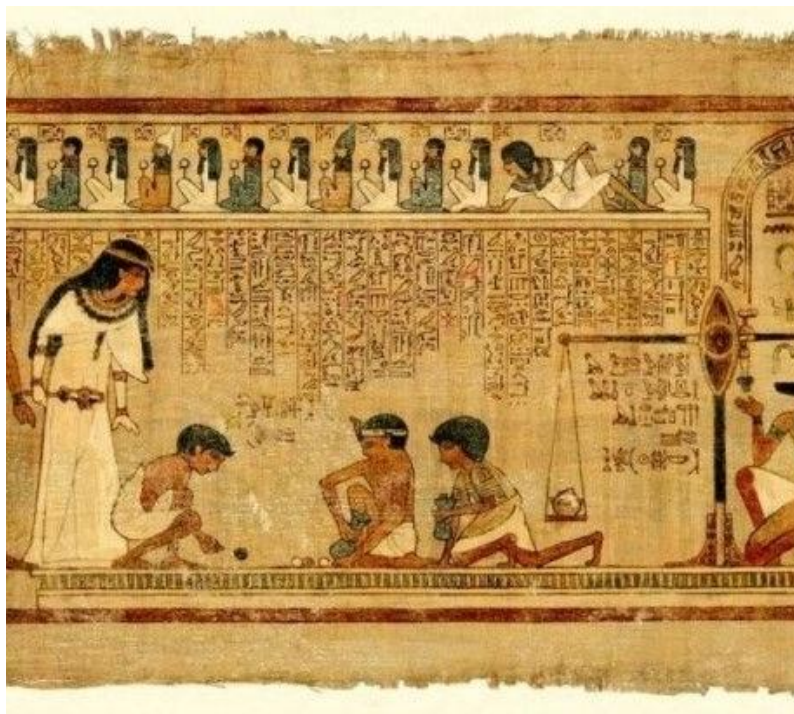
ОТКРЫТЫЕ ВОПРОСЫ



- Что такое базы данных?
- Зачем они нужны?
- Каковы предпосылки к их появлению?



СОХРАНЕНИЕ ИНФОРМАЦИИ



X5 RETAIL GROUP

@ mail.ru
group



Tinkoff
Bank

A



БАЗЫ ДАННЫХ



amazon



ПОВСЮДУ БАЗЫ ДАННЫХ

risovach.ru

Yandex



открытие

БАНК



LEROY MERLIN

БАЗА ДАННЫХ

- совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных
- совокупность данных, организованных в соответствии с концептуальной структурой, описывающей характеристики этих данных и взаимоотношения между ними, причём такое собрание данных, которое поддерживает одну или более областей применения

ЗАЧЕМ НУЖНЫ БАЗЫ ДАННЫХ



ХРАНЕНИЕ И ОБРАБОТКА
БОЛЬШОГО КОЛИЧЕСТВА
ИНФОРМАЦИИ



СОВМЕСТНЫЙ ДОСТУП К
ДАНЫМ

A long time ago in a galaxy far,
far away....

МОДЕЛЬ ДАННЫХ

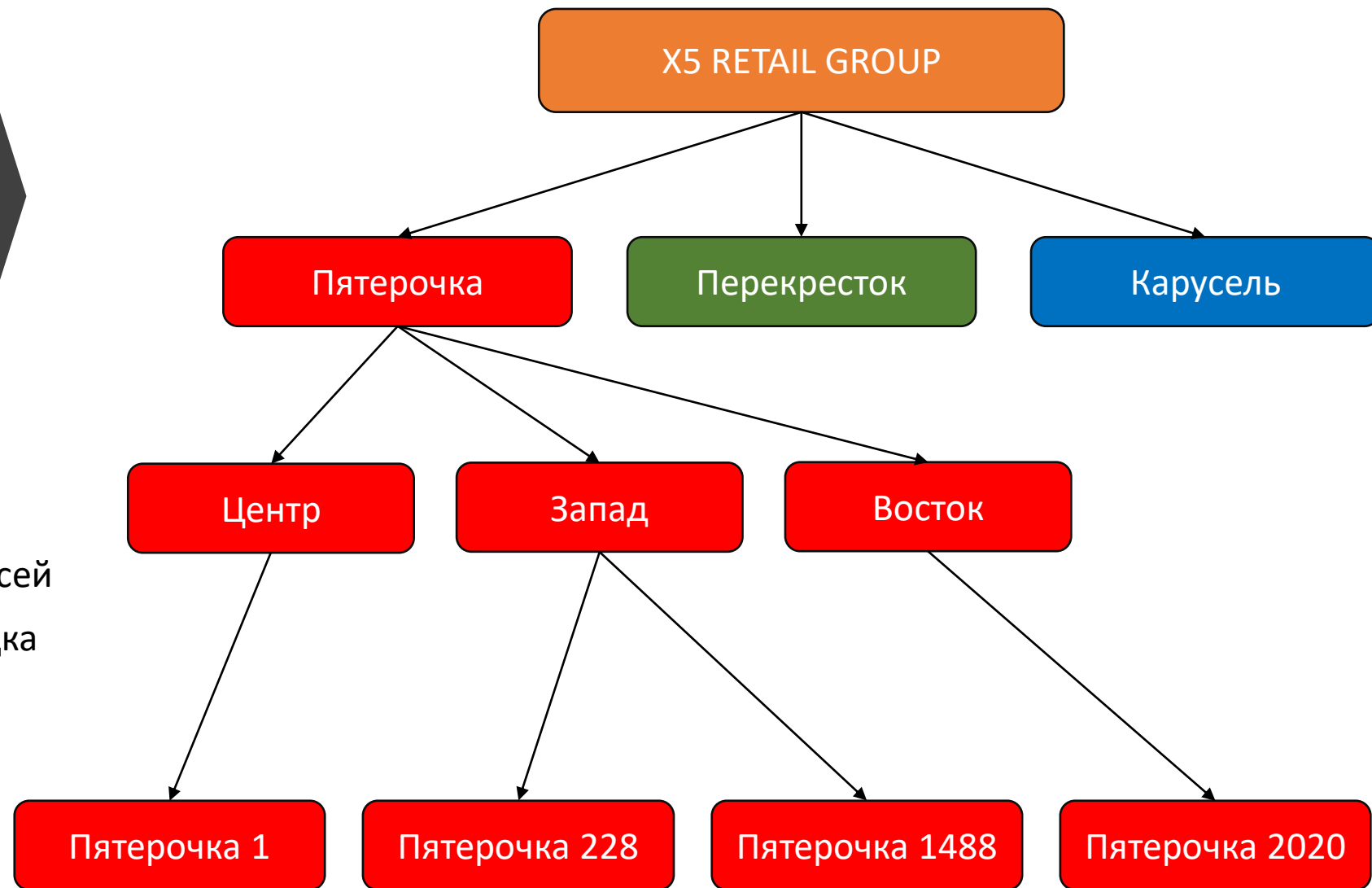
Иерархическая модель данных

Сетевая модель данных

Реляционная модель данных

ИЕРАРХИЧЕСКАЯ МОДЕЛЬ ДАННЫХ

- IBM, 1960s
- Древоподобная структура записей
- Потомок имеет ровно 1 предка
- Потомки общего предка – близнецы



* Все имена вымышлены, все совпадения с реальными компаниями случайны

СЕТЕВАЯ МОДЕЛЬ ДАННЫХ

- Чарльз Бахман, 1969
- Структура записей в виде графа
- Расширяет иерархическую модель данных
- Потомок может иметь более 1 предка



* Все имена вымышлены, все совпадения с реальными компаниями случайны

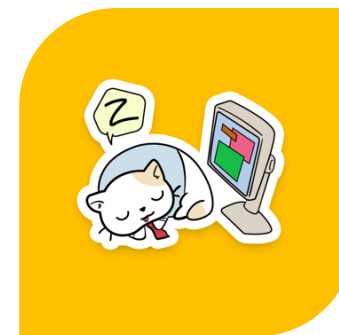
НЕДОСТАТКИ ИЕРАРХИЧЕСКОЙ И СЕТЕВОЙ МОДЕЛЕЙ ДАННЫХ



ВЫСОКАЯ СЛОЖНОСТЬ И
ЖЕСТКОСТЬ СХЕМЫ БД



ОТСУТСТВИЕ ГИБКОСТИ
МОДЕЛИ



ВЫПОЛНЕНИЕ ДАЖЕ
ПРОСТЫХ ЗАПРОСОВ –
СЛОЖНЫЙ ПРОЦЕСС



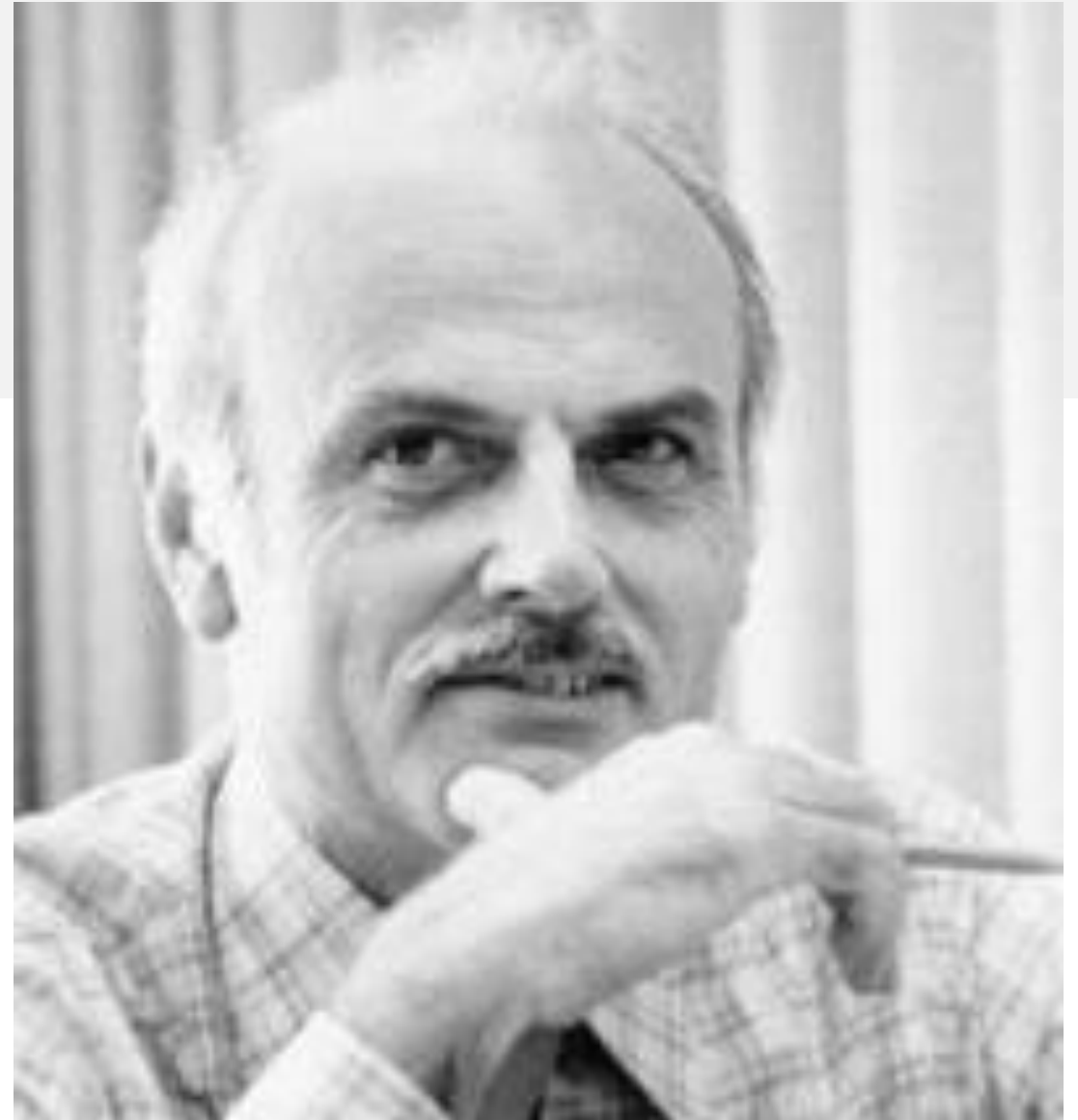
ЗАВИСИМОСТЬ ОТ
ФИЗИЧЕСКОЙ
ОРГАНИЗАЦИИ ДАННЫХ



НИЗКАЯ
ПРОИЗВОДИТЕЛЬНОСТЬ

РЕЛЯЦИОННАЯ МОДЕЛЬ ДАННЫХ

- 1969-1970 гг. Э. Кодд
- Июнь 1970 г. «A Relational Model of Data for Large Shared Data Banks»
- В основе модели – математика и логика
- Логическая модель данных
- Не зависит от физических структур





ОСНОВНЫЕ ПОНЯТИЯ

- *Домен* – это множество допустимых значений
- *Атрибут* – это наименование домена
- *Кортеж* – это упорядоченный набор фиксированной длины
- *Отношение* – это математическая структура, которая формально определяет свойства различных объектов и их взаимосвязи
- *Арность* отношения – количество его элементов

ПРИМЕР РЕЛЯЦИОННОЙ МОДЕЛИ

Тип данных:
integer

Домен:
Идентификатор
человека

Заголовок отношения

ID	FIRST_NM	LAST_NM	PHONE_NO	CREATE_DT
1	Иван	Иванов	79012345678	2017-08-12
101	Сергей	Серов	79023456689	2003-05-14
1006	Петр	Петров	79129876543	2018-06-24
70009	Николай	Сидоров	78123000101	2013-12-16

ПРИМЕР РЕЛЯЦИОННОЙ МОДЕЛИ

Отношение

ID	FIRST_NM	LAST_NM	PHONE_NO	CREATE_DT
1	Иван	Иванов	79012345678	2017-08-12
101	Сергей	Серов	79023456689	2003-05-14
1006	Петр	Петров	79129876543	2018-06-24
70009	Николай	Сидоров	78123000101	2013-12-16

Кортеж

ДЕКАРТОВО ПРОИЗВЕДЕНИЕ

- Пусть A и B – множества:
 - $A = (a_1, a_2, \dots)$
 - $B = (b_1, b_2, \dots)$
- **Декартовым произведением** множеств A и B называется множество пар:

$$A \times B = \{(a, b) : a \in A \text{ \& } b \in B\}$$

РАСШИРЕННОЕ ДЕКАРТОВО ПРОИЗВЕДЕНИЕ

- Пусть S_1, S_2, \dots, S_N - множества, где $N > 1$
 - $S_i = (s_{i1}, s_{i2}, \dots, s_{ij}, \dots)$
- *Расширенное декартово произведение* N множеств – это множество вида:

$$S_1 \times S_2 \times \dots \times S_N = \{(x_1, x_2, \dots, x_N) : x_i \in S_i, i = \overline{1, N}\}$$

- Элемент такого множества (x_1, x_2, \dots, x_N) называется *кортежем*.

СЦЕПЛЕНИЕ КОРТЕЖЕЙ

- Пусть x и y – кортежи:
 - $x = (x_1, x_2, \dots, x_n)$
 - $y = (y_1, y_2, \dots, y_m)$
- Тогда сцеплением кортежей x и y будет кортеж размерности $n + m$ вида:
 - $x \times y = (x_1, \dots, x_n, y_1, \dots, y_m)$



ОСНОВНЫЕ ПОНЯТИЯ

- Пусть заданы домены D_1, D_2, \dots, D_N
- Пусть задан список атрибутов A_1, A_2, \dots, A_N так, что каждому домену D_i соответствует атрибут A_i , определенный на этом домене.
- Тогда *отношением* R , определенным на атрибутах A_1, A_2, \dots, A_N (доменах D_1, D_2, \dots, D_N) называется подмножество расширенного декартова произведения данных доменов:

$$R \subseteq D_1 \times \dots \times D_N$$

РАСШИРЕННОЕ ДЕКАРТОВО ПРОИЗВЕДЕНИЕ

Q

Фамилия	Имя	Отчество
Иванов	Иван	Иванович
Петров	Петр	Петрович
Васильев	Василий	Васильевич

R

Дата визита	Наименование филиала
12.12.2018	«Феникс»
01.01.2018	«Алмаз»

$$Q \times R = Z$$

1. Сколько кортежей будет в новом отношении Z?
2. Сколько атрибутов будет в новом отношении Z?

РАСШИРЕННОЕ ДЕКАРТОВО ПРОИЗВЕДЕНИЕ

Q

Фамилия	Имя	Отчество
Иванов	Иван	Иванович
Петров	Петр	Петрович
Васильев	Василий	Васильевич

R

Дата визита	Наименование филиала
12.12.2018	«Феникс»
01.01.2018	«Алмаз»

$$Q \times R = Z$$

Фамилия	Имя	Отчество	Дата визита	Наименование филиала
Иванов	Иван	Иванович	12.12.2018	«Феникс»
Иванов	Иван	Иванович	01.01.2018	«Алмаз»
Петров	Петр	Петрович	12.12.2018	«Феникс»
Петров	Петр	Петрович	01.01.2018	«Алмаз»
Васильев	Василий	Васильевич	12.12.2018	«Феникс»
Васильев	Василий	Васильевич	01.01.2018	«Алмаз»



ОСНОВНЫЕ ПОНЯТИЯ

- *Арность* отношения – количество атрибутов.
- *Заголовок* отношения – список атрибутов.
- Множество кортежей, входящих в состав отношения, - *тело отношения*.
- Домен называется *составным*, если он представляет собой расширенное декартово произведение конечного числа простых доменов.
- Будем говорить, что два простых домена $D1$ и $D2$ являются *совместимыми*, если они либо совпадают, либо $D2 \subseteq D1$ или $D1 \subseteq D2$.



СВОЙСТВА ОТНОШЕНИЯ

Нет двух одинаковых кортежей

Порядок кортежей не определён

Порядок атрибутов в заголовке не определён

ПРИМЕР РЕЛЯЦИОННОЙ МОДЕЛИ

Отношение. В общем случае таблица отношением не является! Почему?

ID	FIRST_NM	LAST_NM	PHONE_NO	CREATE_DT
1	Иван	Иванов	79012345678	2017-08-12
101	Сергей	Серов	79023456689	2003-05-14
1006	Петр	Петров	79129876543	2018-06-24
70009	Николай	Сидоров	78123000101	2013-12-16

РЕЛЯЦИОННАЯ АЛГЕБРА

Семейство $\mathfrak{A} \subset 2^X$ подмножеств множества X (носитель алгебры) называется *алгеброй*, если оно удовлетворяет следующим свойствам:

- $\emptyset \in \mathfrak{A}$
- Если $A \in \mathfrak{A}$, то $X \setminus A \in \mathfrak{A}$
- Если $A, B \in \mathfrak{A}$, то $A \cup B \in \mathfrak{A}$.

Реляционная алгебра:

- Носитель – множество (всевозможных) отношений различных (конечных) порядков



ОПЕРАЦИИ РЕЛЯЦИОННОЙ АЛГЕБРЫ

Теоретико-множественные

Реляционные

ТЕОРЕТИКО-МНОЖЕСТВЕННЫЕ ОПЕРАЦИИ

Применимы к совместимым отношениям:

- Объединение
- Разность
- Пересечение

Здесь и далее декартово произведение == расширенное декартово произведение, если не оговорено обратное

ОБЪЕДИНЕНИЕ ОТНОШЕНИЙ

Семинаристы ПМИ

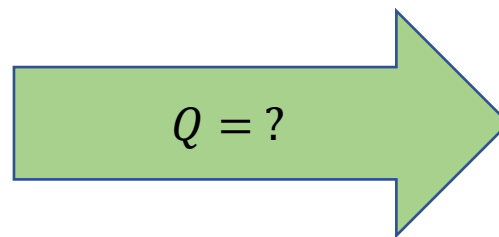
Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

R

Семинаристы не ПМИ

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Роздухова	Нина

S



Все семинаристы



ОБЪЕДИНЕНИЕ ОТНОШЕНИЙ

Семинаристы ПМИ

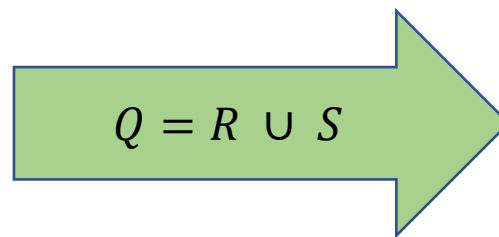
R

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

Семинаристы не ПМИ

S

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Роздухова	Нина



Все семинаристы

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим
Роздухова	Нина

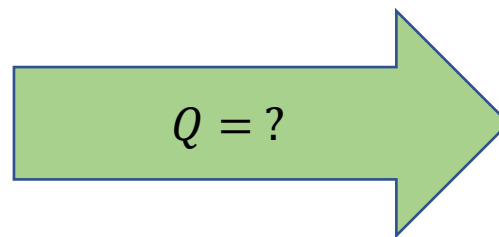
Q

ПЕРЕСЕЧЕНИЕ ОТНОШЕНИЙ

Семинаристы ПМИ

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

R



*Семинаристы, которые преподают
не только на ПМИ*

Семинаристы не ПМИ

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Роздухова	Нина

S

ПЕРЕСЕЧЕНИЕ ОТНОШЕНИЙ

Семинаристы ПМИ

R

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

Семинаристы не ПМИ

S

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Роздухова	Нина


$$Q = R \cap S$$

Семинаристы, которые преподают не только ПМИ

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр

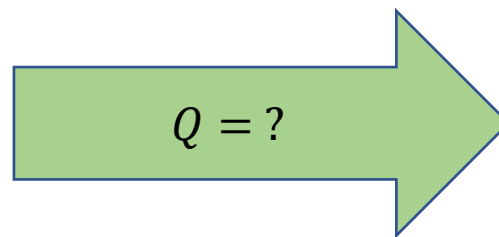
Q

РАЗНОСТЬ ОТНОШЕНИЙ

Семинаристы ПМИ

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

R



Семинаристы, которые
преподают только на ПМИ

Семинаристы не ПМИ

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Роздухова	Нина

S

РАЗНОСТЬ ОТНОШЕНИЙ

Семинаристы ПМИ

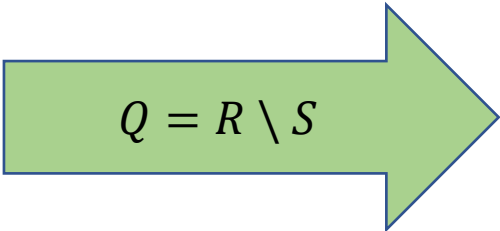
R

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

Семинаристы не ПМИ

S

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Роздухова	Нина


$$Q = R \setminus S$$

Семинаристы, которые преподают только на ПМИ

Фамилия	Имя
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

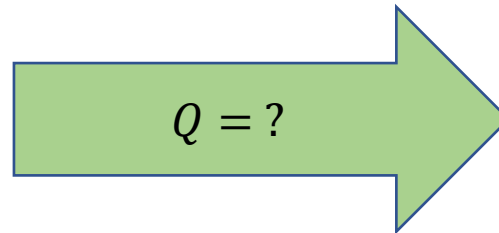
Q

НЕСКОЛЬКО ОПЕРАЦИЙ

Семинаристы ПМИ

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

R



*Семинаристы, которые
преподают ИЛИ только на ПМИ,
ИЛИ только на других
направлениях*

Семинаристы не ПМИ

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Роздухова	Нина

S

НЕСКОЛЬКО ОПЕРАЦИЙ

Семинаристы ПМИ

R

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим

Семинаристы не ПМИ

S

Фамилия	Имя
Меркурьева	Надежда
Халяпов	Александр
Тюрюмина	Элла
Роздухова	Нина

$$Q = (R \setminus S) \cup (S \setminus R)$$

Семинаристы, которые преподают ИЛИ только на ПМИ, ИЛИ только на других направлениях

Фамилия	Имя
Мавлютов	Максим
Лукьянов	Денис
Митюрин	Максим
Тюрюмина	Элла
Роздухова	Нина

Q



РЕЛЯЦИОННЫЕ ОПЕРАЦИИ

Ограничение

Проекция

Соединение

Деление

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ОГРАНИЧЕНИЕ

Построение нового отношения, в которое входят кортежи, удовлетворяющие заданному условию.

- R – заданное отношение,
- A и B – списки идентификаторов атрибутов,
- $\theta \in \{=, \neq, <, >, \geq, \leq\}$.

$$R[A\theta B] = \{r | r \in R \ \& \ (r[A]\theta r[B])\}$$

$r[A], r[B]$ – однотипные

Проблема: допустимо только сравнение между собой значений (составных) атрибутов в рамках одного кортежа

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ОГРАНИЧЕНИЕ

- Пусть α – некоторая константа, совместимая с $r[A]$

$$R[A\theta\alpha] = (R \times _ (B_\alpha)\{(\alpha)\})[A\theta B_\alpha]$$

- $_ (B_\alpha)\{(\alpha)\}$ - безымянное отношение с единственным атрибутом B_α

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ОГРАНИЧЕНИЕ

$$R[\lambda\theta\mu] = \begin{cases} R[A\theta B], \text{ если } \lambda = A, \mu = B \\ R[\alpha\theta B] = (R \times _ (A_\alpha)\{(\alpha)\})[A_\alpha\theta B], \text{ если } \lambda = \alpha, \mu = B \\ R[A\theta\beta] = (R \times _ (B_\beta)\{(\beta)\})[A\theta B_\beta], \text{ если } \lambda = A, \mu = \beta \\ R[\alpha\theta\beta] = (R \times _ (A_\alpha)\{(\alpha)\} \times _ (B_\beta)\{(\beta)\})[A_\alpha\theta B_\beta], \text{ если } \lambda = \alpha, \mu = \beta \end{cases}$$

- A и B – некоторые атрибуты
- α, β – некоторые константы
- Совместимость по типу!

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ОГРАНИЧЕНИЕ

R

person_nm	score_amt
Иван	10
Петр	3
Николай	15
Сергей	20
Илья	0
Анна	5
Максим	30
Дмитрий	7

$Q = R[\text{score_amt} \geq 15]$



*Имена тех, кто набрал не менее
15 баллов за контрольную*

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ОГРАНИЧЕНИЕ

R

person_nm	score_amt
Иван	10
Петр	3
Николай	15
Сергей	20
Илья	0
Анна	5
Максим	30
Дмитрий	7

$Q = R[\text{score_amt} \geq 15]$

Q

person_nm	score_amt
Николай	15
Сергей	20
Максим	30

Имена тех, кто набрал не менее 15 баллов за контрольную

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ПРОЕКЦИЯ

- Построение нового отношения с заданным списком атрибутов.

$$R[L] = \{r[L] | r \in R\}$$

- По Кодду: $r[] = r$, соответственно $R[] = R$.
- Мы будем считать, что результат проекции на пустое множество – отношение с пустым заголовком.

ОТНОШЕНИЕ С ПУСТЫМ ЗАГОЛОВКОМ

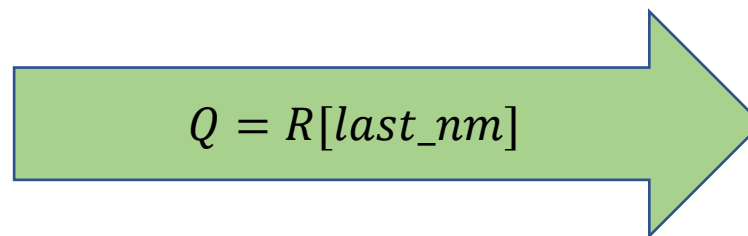
- Существует всего 2:
 - Пустое тело (\perp)
 - Единственный элемент – кортеж нулевой арности (\top)

$$R[] = \begin{cases} \perp, & \text{если } R \text{ было пустым} \\ \top, & \text{если } R \text{ было непустым} \end{cases}$$

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ПРОЕКЦИЯ

R

first_nm	last_nm	score_amt
Иван	Иванов	10
Петр	Петров	3
Николай	Иванов	15
Сергей	Серов	20
Илья	Иванов	0
Анна	Петрова	5
Максим	Серов	30
Дмитрий	Петров	7



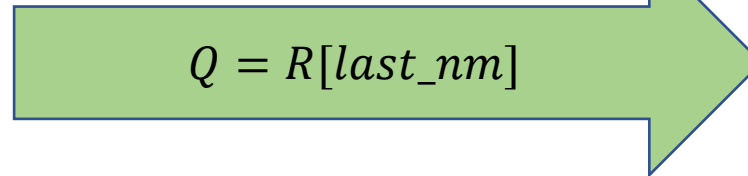
*Уникальные фамилии всех,
кто писал контрольную*



РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ПРОЕКЦИЯ

R

first_nm	last_nm	score_amt
Иван	Иванов	10
Петр	Петров	3
Николай	Иванов	15
Сергей	Серов	20
Илья	Иванов	0
Анна	Петрова	5
Максим	Серов	30
Дмитрий	Петров	7



Q

*Уникальные фамилии всех,
кто писал контрольную*

last_nm
Иванов
Петров
Серов
Петрова

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: СОЕДИНЕНИЕ

- Композиция декартова произведения двух отношений с последующим ограничением по заданному условию.
 - $\theta \in \{=, \neq, <, >, \leq, \geq\}$
 - $R_1(A_1, \dots, A_n)$ — отношение
 - $R_2(B_1, \dots, B_m)$ — отношение

$$R_1 \left[R_1[A_i] \theta R_2[B_j] \right] R_2 = \{r_1 \times r_2 \mid r_1 \in R_1 \ \& \ r_2 \in R_2 \ \& \ r_1[A_i] \theta r_2[B_j]\}$$

- В качестве A_i, B_j можно использовать списки атрибутов.
- *Естественное* соединение: лишний атрибут удаляет

РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: СОЕДИНЕНИЕ

R1

id	uid	task_code	comment_txt
1	123456	A	awful
2	101010	A	excellent
3	123456	B	terrible
4	101010	B	outstanding
5	123456	C	so-so
6	101010	C	unbelievable
7	101010	D	the best of the best

R2

uid	student_nm
123456	You
101010	Son of your mother's friend
136789	Bob

$$Q = R1[R1[UID] = R2[UID]]R2$$

Полная сводка по оценкам
студентов с комментариями



R1 РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: СОЕДИНЕНИЕ

id	uid	task_code	comment_txt
1	123456	A	awful
2	101010	A	excellent
3	123456	B	terrible
4	101010	B	outstanding
5	123456	C	so-so
6	101010	C	unbelievable
7	101010	D	the best of the best

R2

uid	student_nm
123456	You
101010	Son of your mother's friend
136789	Bob

$$Q = R1[R1[UID] = R2[UID]]R2$$

Q

id	uid	task_code	comment_txt	student_nm
1	123456	A	awful	You
2	101010	A	excellent	Son of your mother's friend
3	123456	B	terrible	You
4	101010	B	outstanding	Son of your mother's friend
5	123456	C	so-so	You
6	101010	C	unbelievable	Son of your mother's friend
7	101010	D	the best of the best	Son of your mother's friend

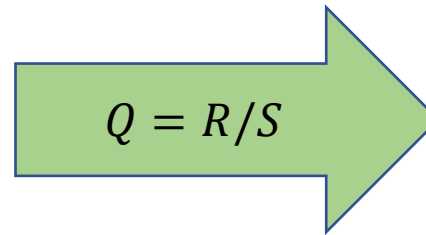
РЕЛЯЦИОННЫЕ ОПЕРАЦИИ: ДЕЛЕНИЕ

R

id	series_nm	channel_nm
0	The Simpsons	RenTV
0	The Simpsons	2x2
0	The Simpsons	CTC
1	Family Guy	RenTV
1	Family Guy	2x2
2	Duck Tales	CTC
2	Duck Tales	2x2

S

channel_nm
RenTV
2x2



Хотим получить сериалы из отношения 1, которые транслировались **по всем** каналам из отношения 2

id	series_nm
0	The Simpsons
1	Family Guy

ЧАТ КУРСА В ТГ

<https://bit.ly/db2022-join>



- Шутить можно

- Обижаться нельзя





ВОПРОСЫ?

