

Нормализация отношений

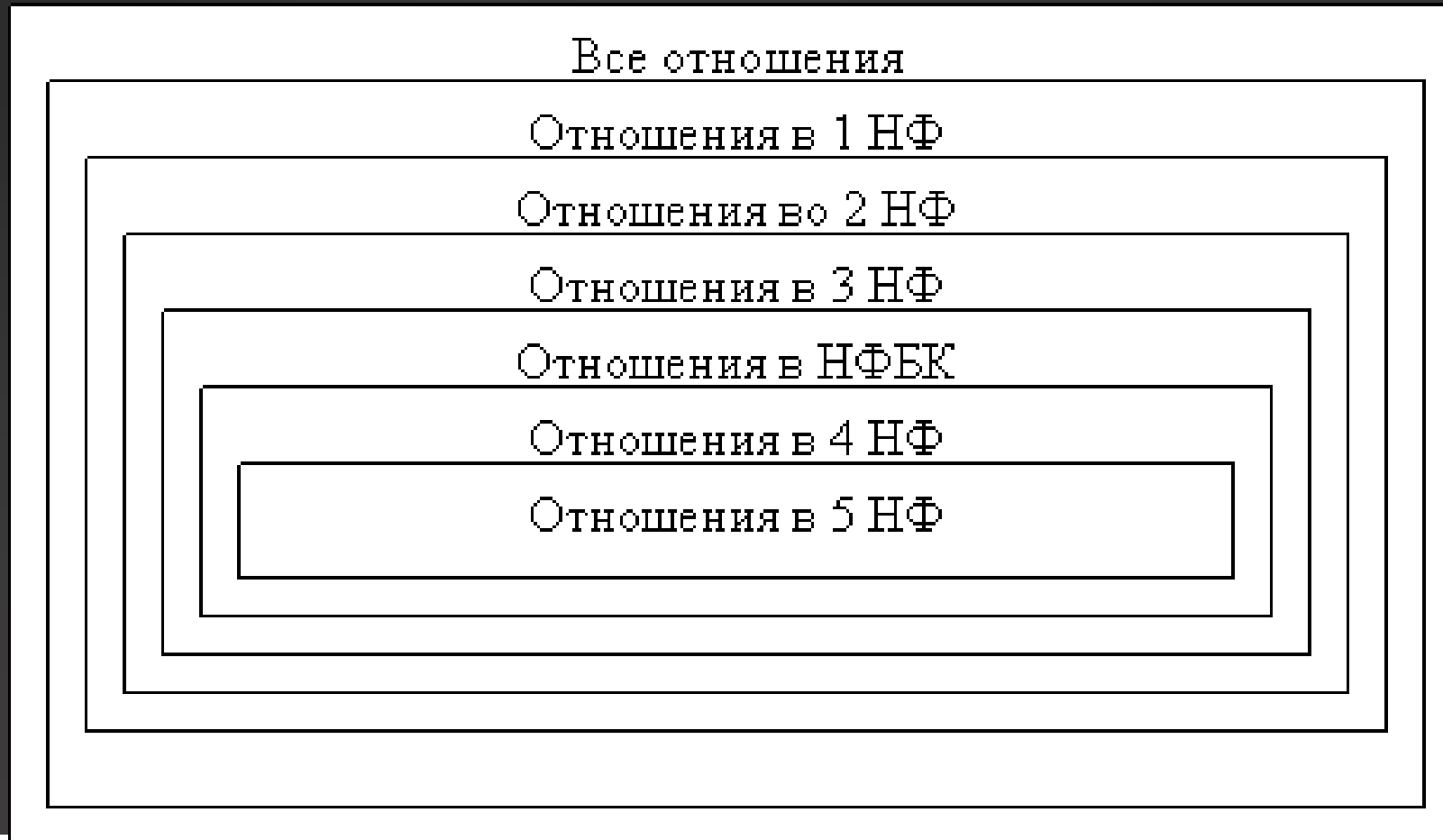


Рис. 4. Нормальные формы

Декомпозиция без потерь

- Процедура нормализации предусматривает разбиение отношения на другие, т.е. его **декомпозицию**. Причем эта декомпозиция должна произойти **без потерь** информации из первоначального отношения. Можно сказать, что декомпозиция должна быть **обратимой**.

Декомпозиция без потерь

Например, рассмотрим отношение
Студенты (№ЗачетнойКнижки, Фамилия, Имя, Отчество,
КодГруппы, Адрес, Телефон).

Выполним декомпозицию на два отношения:

- Студенты1 (№ЗачетнойКнижки, Фамилия, Имя, Отчество) и
- Студенты2 (КодГруппы, Адрес, Телефон)

Декомпозиция без потерь

Выполним декомпозицию по-другому:

- Студенты3 (№ЗачетнойКнижки, Фамилия, Имя, Отчество) и
- Студенты4 (№ЗачетнойКнижки, Адрес, Телефон)

Такое разбиение будет декомпозицией без потерь.

По своей сути декомпозиция – это операция **проекции** реляционной алгебры, поэтому каждое полученное при декомпозиции отношение называют **проекциями** первоначального.

Теорема Хеза

Пусть $R(A, B, C)$ – отношение, где A, B, C – подмножества множества его атрибутов.

Если R удовлетворяет ФЗ $A \rightarrow B$, то R равно соединению его проекций $\{A, B\}$ и $\{A, C\}$.

- (теорема не утверждает «тогда и только тогда»)

Первая НФ

- Отношение находится в **1 НФ** тогда и только тогда, когда все используемые в нем домены содержат только *скалярные значения*.
- Отношение находится в **1 НФ** тогда и только тогда, когда значения всех полей *неделимы*.
- Например, если в отношении есть поле ФИО, но нам необходимо знать отдельно Фамилию, Имя, то отношение не находится в 1 НФ, т.к. значения этого поля можно разделить на фамилию, имя и отчество.

Вторая НФ

- Отношение находится во **2 НФ** тогда и только тогда, когда оно находится в **1 НФ** и каждый неключевой атрибут **неприводимо зависит от первичного ключа**.

Например, рассмотрим отношение Успеваемость (№ЗачетнойКнижки, Фамилия, Имя, Отчество, Дисциплина, Оценка).

Если первичным ключом здесь назначить №ЗачетнойКнижки, то от этого ключа не будет зависеть атрибут Дисциплина. Т.е. в этом случае отношение не находится во 2 НФ.

Вторая НФ

Успеваемость (№ЗачетнойКнижки, Фамилия, Имя, Отчество, Дисциплина, Оценка).

Можно тогда в качестве первичного ключа взять множество атрибутов {№ЗачетнойКнижки, Дисциплина}. В этом случае от такого ключа зависят все атрибуты, но атрибуты Фамилия, Имя, Отчество зависят приводимо:

$\{\text{№ЗачетнойКнижки}, \text{Дисциплина}\} \rightarrow \{\text{Фамилия}, \text{Имя}, \text{Отчество}\}$

И при таком первичном ключе отношение не находится во 2 НФ.

Вторая НФ

Успеваемость (№ЗачетнойКнижки, Фамилия, Имя, Отчество, Дисциплина, Оценка).

$\{\text{№ЗачетнойКнижки}\} \rightarrow \{\text{Фамилия, Имя, Отчество}\} (A \rightarrow B)$

$\{\text{Дисциплина, Оценка}\} (C)$

По теореме Хеза:

- Студенты $\{\text{№ЗачетнойКнижки, Фамилия, Имя, Отчество}\} (\{A, B\})$
- Успеваемость $\{\text{№ЗачетнойКнижки, Дисциплина, Оценка}\} (\{A, C\})$

Третья НФ

- Отношение находится в **3 НФ** тогда и только тогда, когда оно находится во **2 НФ** и каждый неключевой атрибут *нетранзитивно зависит от первичного ключа*.
- Отношение находится в **3 НФ** тогда и только тогда, когда каждый кортеж отношения состоит из значения первичного ключа, которое идентифицирует некоторый объект, и набора *взаимно независимых* (или пустых) значений атрибутов, описывающих этот объект.

Третья НФ

- Например, добавим в отношение Студенты атрибут СтаростаГруппы. Тогда в отношении будут следующие ФЗ:
- $\{\text{№ЗачетнойКнижки}\} \rightarrow \{\text{КодГруппы}\}, \{\text{КодГруппы}\} \rightarrow \{\text{СтаростаГруппы}\}$
- То есть атрибут СтаростаГруппы зависит от первичного ключа (№ЗачетнойКнижки) транзитивно через атрибут КодГруппы, а не напрямую.
- Значит, это отношение не находится в 3 НФ.

Третья НФ

- *1 вариант* основан на ФЗ {№ЗачетнойКнижки}→{Фамилия, Имя, Отчество, КодГруппы}.
- В результате получим такие проекции:
- {№ЗачетнойКнижки, Фамилия, Имя, Отчество, КодГруппы} и
- {№ЗачетнойКнижки, СтаростаГруппы}
- *2 вариант* основан на ФЗ {КодГруппы}→{СтаростаГруппы}.
- В результате получим такие проекции:
- {КодГруппы, СтаростаГруппы}
- {КодГруппы, №ЗачетнойКнижки, Фамилия, Имя, Отчество}

Нормальная форма Бойса-Кодда (НФБК)

Определение 3 НФ не совсем подходит для отношений со следующими **свойствами**:

- отношение имеет два или более потенциальных ключа;
- потенциальные ключи сложные;
- потенциальные ключи перекрываются, т.е. имеют хотя бы один общий атрибут.

Нормальная форма Бойса-Кодда (НФБК)

Отношение находится в **НФБК** тогда и только тогда, когда каждая нетривиальная и неприводимая слева ФЗ обладает *потенциальным ключом в качестве детерминанта*.

Нормальная форма Бойса-Кодда (НФБК)

- Успеваемость (№ЗачетнойКнижки, КодДисциплины, НазваниеДисциплины, Оценка)

Потенциальные ключи: {№ЗачетнойКнижки, КодДисциплины} и {№ЗачетнойКнижки, НазваниеДисциплины}.

Это отношение **не в НФБК**, т.к. в нем есть такие ФЗ:

- КодДисциплины → НазваниеДисциплины
- НазваниеДисциплины → КодДисциплины

А детерминанты этих ФЗ не являются потенциальными ключами.

Нормальная форма Бойса-Кодда (НФБК)

Номер корта	Время начала	Время окончания	Тариф
1	09:30	10:30	«Корт 1 для членов клуба»
1	11:00	12:00	«Корт 1 для членов клуба»
1	14:00	15:30	«Корт 1 для не членов клуба»
2	10:00	11:30	«Корт 2 для не членов клуба»
2	11:30	13:30	«Корт 2 для не членов клуба»
2	15:00	16:30	«Корт 2 для членов клуба»

Возможны следующие составные ключи:

{Номер корта, Время начала}, {Номер корта, Время окончания},
{Тариф, Время начала}, {Тариф, Время окончания}.

- Отношение находится в 3НФ. Требования 2НФ выполняются, так как все атрибуты входят в какой-то из потенциальных ключей, а неключевых атрибутов в отношении нет. Также нет и транзитивных зависимостей, что соответствует требованиям 3НФ. Тем не менее, существует ФЗ Тариф \rightarrow Номер корта, в которой левая часть (детерминант) не является потенциальным ключом отношения, то есть отношение не находится в нормальной форме Бойса — Кодда.
- Недостатком данной структуры является то, что, например, по ошибке можно приписать тариф «Корт 1..» к бронированию корта 2, хотя он может относиться только к корту 1.

Многозначные зависимости

- Пусть A, B, C – произвольные подмножества множества атрибутов отношения R . Тогда B **многозначно зависит** от A тогда и только тогда, когда множество значений B , соответствующее заданной паре значений (значение A , значение C) отношения R , зависит только от A , но не зависит от C .
- Обозначается: $A \twoheadrightarrow B$
- Упрощенно* можно сказать так: A многозначно определяет B , если для каждого значения A не существует **единственного** соответствующего ей значения B (не верна ФЗ $A \rightarrow B$), но каждое значение A определяет **множество** соответствующих ей значений B .

Многозначные зависимости

Дисциплина	Преподаватель	Кабинет
ИС	Огородников Е.В.	140
ИС	Огородников Е.В.	218
ИС	Шиляева М.С.	140
ИС	Шиляева М.С.	218
СУБД	Шиляева М.С.	315
СУБД	Шиляева М.С.	411
СУБД	Петухова М.В.	315
СУБД	Петухова М.В.	411

МЗ:

Дисциплина->>Преподаватель

Дисциплина->>Кабинет

Введем следующие ограничения:

а) каждой дисциплине может соответствовать любое количество преподавателей и любое количество кабинетов;

б) преподаватели и кабинеты не зависят друг от друга (т.е. независимо от того, кто преподает данную дисциплину, для этой дисциплины используется один и тот же набор кабинетов);

с) конкретный преподаватель и конкретный кабинет могут быть связаны с любым количеством дисциплин.

Теорема Фейгина

- Пусть A, B, C – подмножества множества атрибутов отношения R . Отношение R будет равно соединению его проекций $\{A, B\}$ и $\{A, C\}$ тогда и только тогда, когда для отношения R выполняется $M3 A \twoheadrightarrow B | C$.

Четвертая нормальная форма

- Отношение R находится в **4 НФ** тогда и только тогда, когда *если* существуют такие подмножества A и B множества атрибутов R , что выполняется нетривиальная МЗ $A \twoheadrightarrow B$, *то* все атрибуты отношения R функционально зависят от атрибута A .

Или по-другому:

- Отношение R находится в **4 НФ** если оно находится в **НФБК** и все **МЗ** отношения R фактически *являются ФЗ* от потенциальных ключей.

Зависимость соединения

- Если невозможно произвести декомпозицию без потерь одного отношения на две проекции, но возможно разбить отношение на три или более проекций, такое отношение называют **n -декомпозируемым**, где n – число проекций, на которое можно разбить отношение без потерь.

ЛДС (Литература, Дисциплина, Специальность), полностью ключевое.

Литература	Дисциплина	Специальность
Хомоненко А. и др. Базы данных	СУБД	Информатика с доп. спец. ин. язык
Хомоненко А. и др. Базы данных	ИС	Математика с доп. спец. информатика
Хомоненко А. и др. Базы данных	ИС	Информатика с доп. спец. ин. язык
Дейт К. Введение в системы баз данных	СУБД	Информатика с доп. спец. ин. язык
Дейт К. Введение в системы баз данных	ИС	Информатика с доп. спец. ин. язык

Произведем декомпозицию на три проекции: ЛД, ДС, ЛС.

ЛДС (Литература, Дисциплина, Специальность)

Литература	Дисциплина		Дисциплина	Специальность		Литература	Специальность
Хомоненко	СУБД		СУБД	Информатика		Хомоненко	Информатика
Хомоненко	ИС		ИС	Математика		Хомоненко	Математика
Дейт	СУБД		ИС	Информатика		Дейт	Информатика
Дейт	ИС						

ЛДС (Литература, Дисциплина, Специальность)

Соединим обратно только две проекции, ЛД и ДС по атрибуту Дисциплина, и сравним кортежи полученного отношения с кортежами первоначального отношения ЛДС:

Литература	Дисциплина	Специальность	
Хомоненко	СУБД	Информатика	- есть
Хомоненко	ИС	Математика	- есть
Хомоненко	ИС	Информатика	- есть
Дейт	СУБД	Информатика	- есть
Дейт	ИС	Математика	- нет!
Дейт	ИС	Информатика	- есть

Зависимость соединения

- Пусть R – отношение, A, B, \dots, Z – произвольные подмножества множества атрибутов R . Отношение R удовлетворяет зависимости соединения $*(A, B, \dots, Z)$ тогда и только тогда, когда оно равносильно соединению своих проекций с подмножествами атрибутов A, B, \dots, Z .

Для нашего примера отношение ЛДС удовлетворяет зависимости соединения

- $*(\text{ЛД}, \text{ДС}, \text{ЛС})$ или
- $*(\{\text{Литература}, \text{Дисциплина}\}, \{\text{Дисциплина}, \text{Специальность}\}, \{\text{Литература}, \text{Специальность}\})$.

Пятая нормальная форма

- Отношение R находится в 5 НФ, которая также называется **проекционно-соединительной НФ**, тогда и только тогда, когда каждая ЗС в отношении R *подразумевается потенциальными ключами* отношения R .
- Например, рассмотрим отношение Группы (КодГруппы, Курс, Староста). Имеется потенциальный ключ: КодГруппы. Этот ключ подразумевает (т.е. обеспечивает) такие ЗС:
- $\ast(\{\text{КодГруппы, Курс}\} \{\text{КодГруппы, Староста}\})$

Доменно-ключевая нормальная форма

- Переменная отношения находится в ДКНФ тогда и только тогда, когда каждое наложенное на неё ограничение является логическим следствием ограничений доменов и ограничений ключей, наложенных на данную переменную отношения.

Ограничение домена – ограничение, предписывающее использовать для определённого атрибута значения только из некоторого заданного домена. Ограничение по своей сути является заданием перечня (или логического эквивалента перечня) допустимых значений типа и объявлением о том, что указанный атрибут имеет данный тип.

Доменно-ключевая нормальная форма

- Ограничение ключа – ограничение, утверждающее, что некоторый атрибут или комбинация атрибутов является потенциальным ключом.

Любая переменная отношения, находящаяся в ДКНФ, обязательно находится в 5НФ. Однако не любую переменную отношения можно привести к ДКНФ.

Шестая нормальная форма

- Переменная отношения находится в шестой нормальной форме тогда и только тогда, когда она удовлетворяет всем нетривиальным зависимостям соединения. Из определения следует, что переменная находится в 6НФ тогда и только тогда, когда она неприводима, то есть не может быть подвергнута дальнейшей декомпозиции без потерь. Каждая переменная отношения, которая находится в 6НФ, также находится и в 5НФ.

Итоговая схема процедуры нормализации

1. Исключение *избыточности*.
2. Устранение *аномалий* обновления.
3. Проектирование макета данных, который *соответствовал бы реальному миру*, был *интуитивно понятен* и служил основой для *дальнейшего развития*.
4. Упрощение процесса наложения ограничений *целостности*. Эта цель связана с потенциальными ключами, т.е. если соблюдать условие уникальности потенциальных ключей и организовывать связи только через них, то эта цель будет достигнута.