ТЕМА 6. ФУНКЦИОНАЛЬНЫЕ ЗАВИСИМОСТИ

Общие определения

Пусть задана переменная отношения r, и X и Y являются произвольными подмножествами заголовка r ("составными" атрибутами).

В значении переменной отношения r атрибут Y функционально зависит от атрибута X в том и только в том случае, если каждому значению X соответствует в точности одно значение Y. В этом случае говорят также, что атрибут X **функционально определяет** атрибут Y (X является детерминантом (**определителем**) для Y, а Y является зависимым от X). Будем обозначать это как r.X→r.Y.

Для примера будем использовать отношение СЛУЖАЩИЕ_ПРОЕКТЫ {СЛУ_НОМ, СЛУ_ИМЯ, СЛУ_ЗАРП, ПРО_НОМ, ПРОЕКТ_РУК}. Очевидно, что если СЛУ_НОМ является первичным ключом отношения СЛУЖАЩИЕ, то для этого отношения справедлива функциональная зависимость (Functional Dependency – FD) СЛУ_НОМ→СЛУ_ИМЯ.

На самом деле, для тела отношения СЛУЖАЩИЕ_ПРОЕКТЫ в том виде, в котором оно показано на рисунке, выполняются еще и следующие FD:

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП	ПРО_НОМ	ПРОЕКТ_РУК
2934	Иванов	22400.00	1	Иванов
2935	Петров	29600.00	1	Иванов
2936	Сидоров	18000.00	1	Иванов
2937	Федоров	20000.00	1	Иванов
2938	Иванова	22000.00	1	Иванов
2939	Сидоренко	18400.00	2	Иваненко
2940	Федоренко	20400.00	2	Иваненко
2941	Иваненко	22600.00	2	Иваненко

СЛУ_НОМ→СЛУ_ИМЯ

СЛУ_НОМ→СЛУ_ЗАРП

СЛУ_НОМ→ПРО_НОМ

СЛУ_НОМ→ПРОЕКТ_РУК

{СЛУ_НОМ, СЛУ_ИМЯ}→СЛУ_ЗАРП

{СЛУ_НОМ, СЛУ_ИМЯ}→ПРО_НОМ

{СЛУ_НОМ, СЛУ_ИМЯ}→{СЛУ_ЗАРП, ПРО_НОМ}

…

ПРО_НОМ→ПРОЕКТ_РУК и т.д.

Поскольку имена всех служащих различны, то выполняются и такие FD (2):

СЛУ_ИМЯ→СЛУ_НОМ

СЛУ_ИМЯ→СЛУ_ЗАРП

СЛУ_ИМЯ→ПРО_НОМ и т.д.

Более того, для примера на рис. 6.1 выполняется и FD (3):

СЛУ_ЗАРП→ПРО_НОМ

Однако заметим, что природа FD группы (1) отличается от природы FD групп (2) и (3). Логично предположить, что идентификационные номера служащих должны быть всегда различны, а у каждого проекта имеется только один руководитель. Поэтому FD группы (1) должны быть верны для любого допустимого значения переменной отношения СЛУЖАЩИЕ_ПРОЕКТЫ и могут рассматриваться как **инварианты**, или **ограничения целостности** этой переменной отношения.

FD группы (2) базируются на менее естественном предположении о том, что имена всех служащих различны. Это соответствует действительности для примера из рисунка, но возможно, что с течением времени FD группы (2) не будут выполняться для какого-либо значения переменной отношения СЛУЖАЩИЕ_ПРОЕКТЫ.

Наконец, FD группы (3) основана на совсем неестественном предположении, что никакие двое служащих, участвующие в разных проектах, не получают одинаковую зарплату. Опять же, данное предположение верно для примера из рисунка, но, скорее всего, это случайное совпадение.

В дальнейшем нас будут интересовать только те функциональные зависимости, которые должны выполняться для всех возможных значений переменных отношений.

Заметим, что если атрибут A отношения r является возможным ключом, то для любого атрибута B этого отношения всегда выполняется FD $A \rightarrow B$ (в группе (1) к этим FD относятся все FD, детерминантом которых является СЛУ_НОМ). Обратите внимание, что наличие в отношении СЛУЖАЩИЕ_ПРОЕКТЫ FD ПРО_НОМ$\rightarrow$ПРОЕКТ_РУК приводит к некоторой **избыточности** этого отношения. Имя руководителя проекта является характеристикой проекта, а не служащего, но в нашем случае содержится в теле отношения столько раз, сколько служащих работает над проектом.

Итак, мы будем иметь дело с FD, которые выполняются для всех возможных состояний тела соответствующего отношения и могут рассматриваться как ограничения целостности. Как показывает (неполный) список (1), таких зависимостей может быть очень много. Поскольку они трактуются как ограничения целостности, за их соблюдением должна следить СУБД. Поэтому важно уметь сократить набор FD до минимума, поддержка которого гарантирует выполнение всех зависимостей. Мы займемся этим в следующих подразделах.

Замыкание множества функциональных зависимостей. Аксиомы Армстронга. Замыкание множества атрибутов

Замыканием множества FD S является множество FD S^+, включающее все FD, логически выводимые из FD множества S.

Подход к решению проблемы поиска замыкания S^+ множества FD S впервые предложил Вильям Армстронг[1)]. Им был предложен набор правил вывода новых FD из существующих (эти правила обычно называют аксиомами Армстронга, хотя справедливость правил доказывается на основе определения FD). Обычно принято формулировать эти правила вывода в следующей форме. Пусть A, B и C являются (в общем случае, составными) атрибутами отношения r. Множества A, B и C могут иметь непустое пересечение. Для краткости будем обозначать через AB A UNION B. Тогда:

1. если $B \in A$, то $A \rightarrow B$ (рефлексивность);
2. если $A \rightarrow B$, то $AC \rightarrow BC$ (пополнение);
3. если $A \rightarrow B$ и $B \rightarrow C$, то $A \rightarrow C$ (транзитивность).

Можно доказать, что система правил вывода Армстронга **полна** и **совершенна (sound and complete)** в том смысле, что для данного множества FD S любая FD, потенциально выводимая из S, может быть выведена на основе аксиом Армстронга, и применение этих аксиом не может привести к выводу лишней FD. Тем не менее Дейт по практическим соображениям предложил расширить базовый набор правил вывода еще пятью правилами:

4. $A \rightarrow A$ (самодетерминированность) – прямо следует из правила (1);

5. если $A \rightarrow BC$, то $A \rightarrow B$ и $A \rightarrow C$ (декомпозиция) – из правила (1) следует, что $BC \rightarrow B$; по правилу (3) $A \rightarrow B$; аналогично, из $BC \rightarrow C$ и правила (3) следует $A \rightarrow C$;

6. если $A \rightarrow B$ и $A \rightarrow C$, то $A \rightarrow BC$ (объединение) – из правила (2) следует, что $A \rightarrow AB$ и $AB \rightarrow BC$; из правила (3) следует, что $A \rightarrow BC$;

7. если $A \rightarrow B$ и $C \rightarrow D$, то $AC \rightarrow BD$ (композиция) – из правила (2) следует, что $AC \rightarrow BC$ и $BC \rightarrow BD$; из правила (3) следует, что $AC \rightarrow BD$;

8. если $A \rightarrow BC$ и $B \rightarrow D$, то $A \rightarrow BCD$ (накопление) – из правила (2) следует, что $BC \rightarrow BCD$; из правила (3) следует, что $A \rightarrow BCD$.

Минимальное покрытие множества функциональных зависимостей

Множество FD $S2$ называется покрытием множества FD $S1$, если любая FD, выводимая из $S1$, выводится также из $S2$.

Легко заметить, что $S2$ является покрытием $S1$ тогда и только тогда, когда $S1^+ \in S2^+$. Два множества FD $S1$ и $S2$ называются **эквивалентными**, если каждое из них является покрытиемдругого, т. е. $S1^+ = S2^+$.

Множество FD S называется минимальным в том и только в том случае, когда удовлетворяет следующим свойствам:

1. правая часть любой FD из S является множеством из одного атрибута (простым атрибутом);
2. детерминант каждой FD из S обладает свойством **минимальности**; это означает, что удаление любого атрибута из детерминанта приводит к изменению замыкания S^+, т. е. порождению множества FD, не эквивалентного S[3)];
3. удаление любой FD из S приводит к изменению S^+, т. е. порождению множества FD, не эквивалентного S.

Приведем общую схему построения S^- по заданному множеству FD S. Во-первых, используя правило (5) (декомпозиции), мы можем привести множество S к эквивалентному множеству FD $S1$, правые части FD которого содержат только одноэлементные множества (простые атрибуты). Далее, для каждой FD из $S1$, детерминант D $\{D_1, D_2, \ldots, D_n\}$ которой содержит более одного атрибута, будем пытаться удалять атрибуты D_i, получая множество FD $S2$. Если после удаления атрибута D_i $S2$ эквивалентно $S1$, то этот атрибут удаляется, и пробуется следующий атрибут. Назовем $S3$ множество FD, полученное путем допустимого удаления атрибутов из всех детерминантов FD множества $S1$. Наконец, для каждой FD f из множества $S3$ будем проверять эквивалентность множеств $S3$ и $S3$ MINUS $\{f\}$. Если эти множества эквивалентны, удалим f из множества $S3$, и в заключение получим множество $S4$, которое минимально и эквивалентно исходному множеству FD S.

Пусть, например, имеется отношение r $\{A, B, C, D\}$ и задано множество FD $S = \{A \rightarrow B, A \rightarrow BC, AB \rightarrow C, AC \rightarrow D, B \rightarrow C\}$. По правилу декомпозиции S эквивалентно множеству $S1$ $\{A \rightarrow B, A \rightarrow C, AB \rightarrow C, AC \rightarrow D, B \rightarrow C\}$. В детерминанте FD $AC \rightarrow D$ можно удалить атрибут C, поскольку по правилу дополнения из FD $A \rightarrow C$ следует $A \rightarrow AC$; по правилу транзитивности выводится FD $A \rightarrow D$, поэтому атрибут C в детерминанте FD $AC \rightarrow D$ является избыточным. FD $AB \rightarrow C$ может быть удалена, поскольку может быть выведена из FD $A \rightarrow C$ (по правилу пополнения из этой FD выводится $AB \rightarrow BC$, а по правилу декомпозиции далее выводится $AB \rightarrow C$). Наконец, FD $A \rightarrow C$ тоже выводится по правилу транзитивности из FD $A \rightarrow B$ и $B \rightarrow C$. Таким образом, мы получаем множество зависимостей $\{A \rightarrow B, A \rightarrow D, B \rightarrow C\}$, которое является минимальным и эквивалентно S по построению.

Декомпозиция без потерь и функциональные зависимости

Считаются правильными такие декомпозиции отношения, которые обратимы, т. е. имеется возможность собрать исходное отношение из декомпозированных отношений без потери информации. Такие декомпозиции называются декомпозициями без потерь.

На рисунке приведены две возможные декомпозиции отношения СЛУЖАЩИЕ_ПРОЕКТЫ.

СЛУЖАЩИЕ_ПРОЕКТЫ

СЛУ_НОМЕР	СЛУ_ИМЯ	СЛУ_ЗАРП	ПРО_НОМ	ПРОЕКТ_РУК
2934	Иванов	22000.00	1	Иванов
2941	Иваненко	22000.00	2	Иваненко

Декомпозиция (1). Отношения **СЛУЖ** и **СЛУ_ПРО**

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП
2934	Иванов	22000.00
2941	Иваненко	22000.00

СЛУ_НОМ	ПРО_НОМ	ПРОЕКТ_РУК
2934	1	Иванов
2941	2	Иваненко

Декомпозиция (2). Отношения **СЛУЖ** и **ЗАРП_ПРО**

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП
2934	Иванов	22000.00
2941	Иваненко	22000.00

СЛУ_ЗАРП	ПРО_НОМ	ПРОЕКТ_РУК
22000.00	1	Иванов
22000.00	2	Иваненко

Анализ показывает, что в случае декомпозиции (1) мы не потеряли информацию о служащих – про каждого из них можно узнать имя, размер зарплаты, номер выполняемого проекта и имя руководителя проекта. Вторая декомпозиция не дает возможности получить данные о проекте служащего, поскольку Иванов и Иваненко получают одинаковую зарплату, следовательно, эта декомпозиция приводит к потере информации.

Корректность же декомпозиции 1 следует из теоремы Хита:

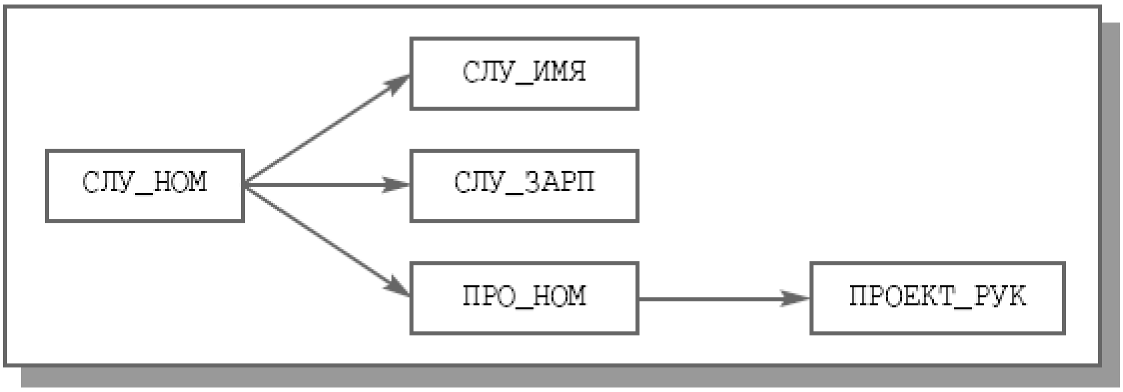
Теорема Хита.

Пусть задано отношение r {A, B, C} (A, B и C, в общем случае, являются составными атрибутами) и выполняется FD A$\rightarrow$B.

СЛУ_НОМ	СЛУ_ИМЯ	СЛУ_ЗАРП	ПРО_НОМ	ПРОЕКТ_РУК
2934	Иванов	22000.00	1	Иванов
2941	Иваненко	22000.00	2	Иваненко
2934	Иванов	22000.00	2	Иваненко
2941	Иваненко	22000.00	1	Иванов

Тогда r = (r PROJECT {A, B}) NATURAL JOIN (r PROJECT {A, C}).

Для наглядной иллюстрации используются диаграммы FD, с помощью которых можно наглядно представлять минимальные множества FD. Например, на рисунке приведена диаграмма минимального множества FD отношения СЛУЖАЩИЕ_ПРОЕКТЫ.



В левой части диаграммы все стрелки начинаются с атрибута СЛУ_НОМ, который является единственным возможным (и, следовательно, первичным) ключом отношения СЛУЖАЩИЕ_ПРОЕКТЫ. Обратите внимание на отсутствие стрелки от СЛУ_НОМ к ПРОЕКТ_РУК. Конечно, поскольку СЛУ_НОМ является возможным ключом, должна выполняться и FD СЛУ_НОМ→ПРОЕКТ_РУК. Но эта FDявляется транзитивной (через ПРО_НОМ) и поэтому не входит в минимальное множество FD. Заметим, что в процессе нормализации, к рассмотрению которого мы приступим в следующей лекции, из диаграмм множества FD удаляются стрелки, начинающиеся не от возможных ключей.