

Тема №3 Функциональные зависимости

1. Понятие функциональной зависимости в данных
2. Основные классы функциональных зависимостей
3. Аксиомы вывода функциональных зависимостей

Понятие функциональной зависимости в данных

Основной единицей представления данных в реляционной модели является отношение, которое математически задается списком имен атрибутов, иначе - схемой отношения.

На стадии логического проектирования реляционной базы данных проектировщик определяет схемы отношений в рамках некоторой предметной области: представляет сущности, группирует их атрибуты, выявляет основные связи между сущностями.

Проектирование реляционной базы данных заключается в обоснованном выборе конкретных схем отношений из множества различных альтернативных вариантов схем.

Понятие функциональной зависимости в данных

На практике построение логической модели базы данных, независимо от используемой модели данных, выполняется с учетом *двух основных требований*:

- **исключить избыточность и**
- **максимально повысить надежность данных.**

Эти требования вытекают из требования коллективного использования данных группой пользователей.

Понятие функциональной зависимости в данных

Формальных средств описания данных, необходимых для проверки правильности заполнения конструкций моделей, недостаточно.

Выбор сущностей, атрибутов и фиксация взаимосвязей между сущностями зависит от семантики предметной области и выполняется системным аналитиком субъективно в соответствии с его личным пониманием специфики прикладной задачи.

Разные люди определяют и представляют данные по-разному!

Понятие функциональной зависимости в данных

Любое априорное знание об ограничениях предметной области, накладываемых на взаимосвязи между данными и значения данных, и знания об их свойствах и взаимоотношениях между ними может сыграть важную роль.

Формализация таких априорных знаний о свойствах данных предметной области базы данных нашла свое отражение в концепции *функциональной зависимости* данных, т.е. ограничений на возможные взаимосвязи между данными, которые могут быть текущими значениями схемы отношений.

Понятие функциональной зависимости в данных

Кортежи отношений могут представлять экземпляры сущности предметной области или фиксировать их взаимосвязь.

Но даже если эти кортежи определены правильно, т.е. отвечают схеме отношения и выбраны из допустимых доменов, *не всякий из них может быть текущим значением некоторого отношения.*

Пример: возраст человека редко бывает более 120 лет, или один и тот же пилот не может одновременно выполнять два различных рейса.

Понятие функциональной зависимости в данных

Определение функции не накладывает никаких ограничений на множество аргументов и множество значений функции, кроме их существования и наличия соответствия между их элементами.

Поскольку ФЗ можно задать таблично, а таблица есть форма представления отношения, то становится очевидной связь между ФЗ и отношением.

Отношение может задавать ФЗ.

Понятие функциональной зависимости в данных

Определение 1.

Пусть $r(A_1, A_2, \dots, A_n)$ - схема отношения R , а X и Y - подмножества r . Говорят, что X функционально определяет Y , если каждому значению атрибутов кортежа отношения из X соответствует не более одного значения атрибутов того же кортежа отношения из Y . Такая ФЗ обозначается как

$$F : X \rightarrow Y$$

Понятие функциональной зависимости в данных

Пример:

ГРАФИК_ПОЛЕТОВ (Пилот, Рейс, Дата_вылета, Время_вылета)

Иванов	100	8.07	10:20
Иванов	102	9.07	13:30
Исаев	90	7.07	6:00
Исаев	100	11.07	10:20
Исаев	103	10.07	19:30
Петров	100	12.07	10:20
Петров	102	11.07	13:30
Фролов	90	8.07	6:00
Фролов	90	12.07	6:00
Фролов	104	14.07	13:30

Понятие функциональной зависимости в данных

Известно, что:

- каждому рейсу соответствует определенное время вылета;
- для каждого пилота, даты и времени вылета возможен только один рейс;
- на определенный день и рейс назначается определенный пилот.

Следовательно:

"Время_вылета" функционально зависим от "Рейс":

"Рейс" → "Время_вылета";

"Рейс" функционально зависим от {"Пилот", "Дата_вылета", "Время_вылета"}:

{"Пилот", "Дата_вылета", "Время_вылета"} → "Рейс";

"Пилот" функционально зависим от {"Рейс", "Дата_вылета"}:

{"Рейс", "Дата_вылета"} → "Пилот".

Понятие функциональной зависимости в данных

Важной задачей при выявлении *функциональных зависимостей* на атрибутах отношения, которое по определению является множеством, является выяснение, какой из атрибутов выступает как аргумент, а какой - как значение ФЗ.

Наиболее подходящими кандидатами в аргументы ФЗ являются *возможные ключи*, так как кортежи представляют экземпляры сущности, которые идентифицируются значениями атрибутов своего ключа.

Нестрого говоря, функциональная зависимость имеет место на отношении, когда значения кортежа на одном множестве атрибутов однозначным образом определяют значения кортежа на другом множестве атрибутов.

Понятие функциональной зависимости в данных

Для получения формального (строгого) определения наличия ФЗ в отношении нужно обратиться к реляционным операциям.

Определение 2.

Пусть имеется отношение R со схемой r , X и Y - два подмножества R . ФЗ $F : X \rightarrow Y$ имеет место на R , если множество $\pi_Y(\sigma_{X=x}(R))$ имеет не более одного кортежа для каждого значения x .

Такая ФЗ называется также F -зависимостью.

Понятие функциональной зависимости в данных

Если семантика предметной области базы данных сложна, то проверить кортежи на принадлежность к ФЗ достаточно сложно.

Сложно вообще установить наличие самой функциональной зависимости, отвечающей природе рассматриваемых данных.

С помощью такого формального метода можно выявить ФЗ, которые не являются реальными и носят случайный характер. Проектировщику реляционных баз данных следует знать о таком методе проверки наличия ФЗ, но при проектировании новой базы данных его применение малоэффективно. *Он может быть полезен при реинжиниринге существующей базы данных.*

Основные классы функциональных зависимостей

Анализ связей между сущностями в предметных областях позволяет выделить различные классы *функциональных зависимостей*.

Значения атрибутов могут зависеть от ключа по-разному. Различают классы *полных* и *частичных ФЗ*.

ФЗ может быть частичной, когда значение неключевого атрибута зависит от значений некоторых атрибутов составного ключа, и полной, когда значения неключевого атрибута зависят от значений всех атрибутов составного ключа.

Основные классы функциональных зависимостей

Определение 3.

Говорят, что неключевой атрибут функционально полно зависит от составного ключа, если он функционально зависит от ключа, но не находится в функциональной зависимости ни от какой части составного ключа.

Если неключевой атрибут зависит от части составного ключа, то говорят о частичной ФЗ.

Основные классы функциональных зависимостей

Пример: Частичные и полные ФЗ

ПРЕПОДАВАТЕЛЬ_ПРЕДМЕТ (Личный номер, Предмет, Фамилия, Должность, Оклад, Часы)

1.	Иванов	доцент	25000	Математика	40
2.	Исаев	доцент	25000	Физика	50
3.	Фролов	профессор	50000	Химия	30

Основные классы функциональных зависимостей

Первичным ключом отношения

ПРЕПОДАВАТЕЛЬ_ПРЕДМЕТ является пара атрибутов

Личный_номер-Предмет. Значения атрибута

Количество_часов зависят от значения атрибута Предмет, т.

е. имеем частичную ФЗ

Предмет → Часы.

Значения атрибута Фамилия зависят от значений атрибутов

Личный_номер-Предмет, т.е. имеем полную

функциональную зависимость

{Личный_номер, Предмет} → Фамилия.

Основные классы функциональных зависимостей

Рассмотрим проблему избыточности данных с точки зрения существования определенных функциональных зависимостей.

Избыточность данных может проявляться в виде дублирования значений некоторых атрибутов.

Так, например, если несколько преподавателей находятся на одной и той же должности, то их оклады могут совпадать.

Атрибут Оклад однозначно определяется атрибутом Должность. Разделение исходного отношения на два новых отношения позволит исключить дублирование данных.

Основные классы функциональных зависимостей

Таким образом, выявление определенных *функциональных зависимостей* в отношениях базы данных позволяет преобразовать их с целью исключения избыточности и повышения надежности данных.

Формирование схем отношений путем разбиения исходных отношений по их атрибутам с учетом функциональных зависимостей является одним из способов создания хороших схем реляционных баз данных.

Основные классы функциональных зависимостей

Каким образом можно использовать это наблюдение с учетом семантики данных для конструирования отношений?

Имеет смысл разбить все возможные зависимости на определенные типы ФЗ, и на основе этой классификации проанализировать, какие типы ФЗ к каким аномалиям в выполнении реляционных операций приводят.

Основные классы функциональных зависимостей

Анализ связей между сущностями в предметных областях позволяет определить, наряду с *частичной* и *полной* ФЗ, еще несколько классов ФЗ.

Одним из таких классов является класс *транзитивных* ФЗ.

Определение 4.

Пусть X, Y, Z - атрибуты отношения R . Если при этом имеются ФЗ $X \rightarrow Y$ и $Y \rightarrow Z$, но отсутствуют ФЗ $Y \rightarrow X$ и $Z \rightarrow Y$ то говорят, что Z транзитивно зависит от X . Такие ФЗ называются транзитивными (Т-зависимостями).

Основные классы функциональных зависимостей

Пример: Транзитивные ФЗ

Личный номер преподавателя определяет его должность, т.е. имеет место ФЗ Личный_номер→Должность.

С другой стороны, согласно тарификации каждой должности назначается определенный оклад, т.е. имеет место ФЗ Должность→Оклад.

Каждый преподаватель получает за работу соответствующий должности оклад, т.е. оклад преподавателя определяется через его должность.

Основные классы функциональных зависимостей

Семантическая связь между атрибутами отношения может носить неоднозначный характер, это порождает существование класса *многозначных зависимостей* (MV-зависимостей).

Пример: один преподаватель может преподавать несколько предметов, а один предмет может преподаваться несколькими преподавателями.

Многозначная зависимость может быть следующих типов: 1:N (один ко многим), M:1 (многие к одному) и M:N (многие ко многим).

Основные классы функциональных зависимостей

Определение 5.

Пусть r - некоторая схема отношения, X и Y - подмножества атрибутов r . Если при заданных значениях атрибутов из $\{X\}$ существует некоторое множество, состоящее из нуля или более взаимосвязанных значений атрибутов из $\{Y\}$, никак не связанных со значениями других атрибутов этого отношения r - X - Y , то говорят о существовании многозначной зависимости между атрибутами X и Y : (класс MV -зависимостей). $MV : X \twoheadrightarrow Y$

Основные классы функциональных зависимостей

Разделение установленных функциональных зависимостей по различным отношениям может привести к *нарушению принципа замкнутости реляционных операций, потере некоторых существующих кортежей* или *появлению мнимых кортежей*.

Поэтому есть необходимость выделения еще одного класса функциональных зависимостей - класса зависимостей по соединению (J-зависимостей).

Этот класс ФЗ требует от ФЗ наличия свойства восстанавливаемости по своим проекциям с помощью естественного соединения.

Основные классы функциональных зависимостей

Пусть U - универсальное отношение, полученное объединением всех атрибутов сущностей предметной области в одно отношение.

Определение 6.

Пусть $r = \{r_1, \dots, r_p\}$ - множество схем на U .
Отношение R из множества U удовлетворяет зависимости по соединению, если R разлагается без потерь на r как

$$R = \pi_{r_1}(R) \bowtie \pi_{r_2}(R) \bowtie \dots \bowtie \pi_p(R)$$

Аксиомы вывода функциональных зависимостей

Для каждой базы данных на множестве ее отношений можно рассмотреть все возможные, допустимые в семантическом смысле функциональные зависимости.

Для каждого отношения существует вполне определенное множество ФЗ между его атрибутами.

На практике число рассматриваемых атрибутов и ФЗ конечно (!).

Аксиомы вывода функциональных зависимостей

Поскольку ФЗ являются высказываниями об атрибутах сущностей предметной области, то над ними могут быть определены операции, позволяющие логически получать одну зависимость из другой (или устанавливать между ними эквивалентность).

Это позволяет определить для данной схемы базы данных базовый набор ФЗ, из которого может быть выведено все множество ФЗ, присущих этой схеме.

Данное утверждение является важной конструктивной идеей в теории проектирования реляционных баз данных.

Аксиомы вывода функциональных зависимостей

Математически задачу вывода базового набора ФЗ можно поставить следующим образом.

Пусть $U = \{A_1, A_2, \dots, A_n\}$ - универсальное множество атрибутов, т.е. полный набор атрибутов отношения базы данных. Совокупность пар (X, Y) , таких, что $X \rightarrow Y$, задает структуру ФЗ отношения R . Такое отношение называют еще *универсальным отношением*.

Задача состоит в построении такого набора ФЗ, из которого могут быть получены все ФЗ базы данных.

Аксиомы вывода функциональных зависимостей

Определение 7.

Пусть F - множество ФЗ для схемы отношения r , - $X \rightarrow Y$ некоторая ФЗ. Говорят, что ФЗ $X \rightarrow Y$ логически следует из F , если для каждого отношения R со схемой r , удовлетворяющего ФЗ из F , удовлетворяется также зависимость $X \rightarrow Y$