

Нормални форми(тема 24)

Обикновено първо проектираме E/R модел на БД, на базата на който определяме същността на данните и връзките между обектите, които ще съхраняваме, след което правим преход към релационното представяне на данните, като се опитваме да премахнем недостатъци и аномалии и на двете нива.

По-конкретно, релационните схеми могат да бъдат подобрени чрез анализиране на функционалните и многозначните функционални зависимости и ограничения.

Аномалии

Излишествата в една релация, т.е. записването на едни и същи данни на различни места, може да доведе до няколко вида неконсистентности или аномалии:

- при неуспешно **обновяване** на всички срещания на данните
- при изтриване
 - например в таблицата по-долу, премахвайки информация за филма Mighty Ducks, премахваме информация и за актьора Emilio Estevez
- при добавяне
 - например в таблицата по-долу не можем да въведем информация за филм без да въведем информация за актьор
- **Movies** релацията - тук за всяко участие на актьор в даден филм за филма се повтаря информация за дължина, тип и име на студио

title	year	length	filmType	studioName	starName
Star Wars	1977	124	color	Fox	Carry Fisher
Mighty Ducks	1991	104	color	Disney	Emilio Estevez
Wayne's World	1992	95	color	Paramount	Dana Carvey
Star Wars	1977	124	color	Fox	Mark Hamill
Star Wars	1977	124	color	Fox	Harrison Ford
Wayne's World	1992	95	color	Paramount	Mike Meyers

Ограничения

- правила, които трябва да важат за всеки екземпляр на дадена релация
- релацията се изменя във времето, но правилата трябва да са в сила
- има различни видове ограничения
 - тези, които се налагат от домейните на атрибутите
 - също напр. UNIQUE ограничението, което забранява два кортежа да имат една и съща стойност за атрибута, за който е приложено
 - ограничения, които се налагат от функционалните зависимости

Функционални зависимости, аксиоми на Армстронг

Функционални зависимости

def Нека R е релация със схема $R(A_1, \dots, A_n, B_1, \dots, B_m, C_1, \dots, C_k)$
Казваме, че множеството атрибути $A := \{A_1, \dots, A_n\}$ **функционално определя** множеството атрибути $B := \{B_1, \dots, B_m\}$ и

пишем $A \rightarrow B$ тогава и само тогава, когато за произволен екземпляр на релацията $r(R)$ е в сила, че $(\forall t, u \in r)[t[A] = u[A] \rightarrow t[B] = u[B]]$.

забл. ФЗ е твърдение за схемата на релацията, т.е. за абстрактната релация и следователно е ограничение, което трябва да бъде изпълнено за всеки екземпляр на релацията

def Нека R е релация със схема $R(A)$, където A е множеството от атрибутите ѝ, $X \subseteq A$. Функционалната зависимост $X \rightarrow Y$ наричаме **тривиална** $\iff Y \subseteq X$.

def Една FD наричаме нетривиална тогава и само тогава, когато не е тривиална.

def Нека R е релация със схема $R(A)$ и $K \subseteq A$. Казваме, че множеството атрибути K е **суперключ** за релацията R тогава и само тогава, когато $K \rightarrow A$, т.е. функционално определя множеството от всички атрибути на релацията.

def Нека R е релация със схема $R(A)$ и $K \subseteq A$. Казваме, че множеството атрибути K е **ключ** за релацията R тогава и само тогава, когато K е суперключ и е в сила $(\forall S \subseteq K)[\text{Se суперключ} \rightarrow S = K]$.

Т.е. ключовете представляват минимални суперключове.

Аксиоми на Армстронг

Аксиомите на Армстронг са правила, отнасящи се за ФЗ, и се наричат аксиоми, понеже чрез тях и базовите функционални зависимости, зададени върху релационната схема, можем да получим всички функционални зависимости, които логически следват.

В този смисъл, те са пълни и надеждни (complete and consistent).

За следващите формулировки, нека R е релация с релационна схема $R(A)$, където A е множеството от атрибутите ѝ.

Нека също $r(R)$ да бъде произволен екземпляр на релацията.

! забл. Доказателства не се изискват в темите.

1. Рефлексивност (i)

Нека $X \subseteq A, Y \subseteq A$ са множества от атрибути. Нека също $Y \subseteq X$. Тогава е в сила, че $X \rightarrow Y$.

Това е така, тъй като ако $Y \subseteq X$ и $t_1, t_2 \in r$ са произволни кортежи от релацията, то тривиално, ако $t_1[X] = t_2[X]$, имаме, че $t_1[Y] = t_2[Y]$. Следователно $X \rightarrow Y$.

2. Разширение (ii)

Нека $X \subseteq A, Y \subseteq A, W \subseteq A$ са множества от атрибути и е в сила, че $X \rightarrow Y$. Тогава е в сила, че $XW \rightarrow YW$.

Това е така, тъй като ако $X \rightarrow Y$ и $t_1, t_2 \in r$ са такива, че за $t_1[XW] = t_2[XW]$, то $t_1[X] = t_2[X]$ и $t_1[W] = t_2[W]$.

Така имаме, че $t_1[Y] = t_2[Y]$ и $t_1[W] = t_2[W]$, следователно $t_1[YW] = t_2[YW]$.

3. Транзитивност (iii)

Нека $X \subseteq A, Y \subseteq A, W \subseteq A$ са множества от атрибути и е в сила, че $X \rightarrow Y$ и $Y \rightarrow W$. Тогава е в сила, че $X \rightarrow W$.

Това е така, тъй като ако $X \rightarrow Y, Y \rightarrow W$ и $t_1, t_2 \in r$ са такива, че за $t_1[X] = t_2[X]$, то $t_1[Y] = t_2[Y]$. Тогава $t_1[W] = t_2[W]$, следователно $X \rightarrow W$.

От аксиомите на Армстронг има няколко по-често използвани следствия:

1. Обединение

Нека $X \subseteq A, Y \subseteq A, W \subseteq A$ са множества от атрибути и е в сила, че $X \rightarrow Y$ и $X \rightarrow W$. Тогава е в сила, че $X \rightarrow YW$.

Това е така, т.к. $X \rightarrow Y$, следователно от (ii) $X \rightarrow XY$; и т.к. $X \rightarrow W$, от (ii) имаме, че $XY \rightarrow YW$. Следователно от (iii) получаваме $X \rightarrow YW$.

2. Псевдотранзитивност

Нека $X \subseteq A, Y \subseteq A, Z \subseteq A$ са множества от атрибути и е в сила, че $X \rightarrow Y$ и $YW \rightarrow Z$. Тогава е в сила, че $XW \rightarrow Z$.

Това е така, т.к. $X \rightarrow Y$, следователно от (ii) $XW \rightarrow YW$ и сега от (iii) получаваме $XW \rightarrow Z$.

3. Декомпозиция(на ФЗ)

Нека $X \subseteq A, Y \subseteq A, Z \subseteq Y$ са множества от атрибути и е в сила, че $X \rightarrow Y$. Тогава е в сила, че $X \rightarrow Z$.

Това е така, т.к. $Z \subseteq Y$ и по (i) имаме $Y \rightarrow Z$, така по (iii) $X \rightarrow Z$.

Ясно е, че можем да разделим дясната част на ФЗ и чрез правилото за обединение да възстановим първоначалната зависимост, т.е.

ако имаме $A \rightarrow \{B_1, \dots, B_m\}$, то

$A \rightarrow B_1,$

$A \rightarrow B_2,$

...,

$A \rightarrow B_m$

Да **забележим**, че не може в общия случай да разделяме лявата част на ФЗ и резултатните ФЗ да бъдат коректни.

4. Правило на тривиалната зависимост

Нека $AA := \{A_1, \dots, A_n\} \subseteq A, BB := \{B_1, \dots, B_m\} \subseteq A, C := AA \cap BB$ и $AA \rightarrow BB$

Тогава $AA \rightarrow BB \setminus C$

Декомпозиция(на релации)

Нека R е релация със схема $R(A_0)$. Тогава $R_1(A_1), \dots, R_n(A_n)$ релациите представляват декомпозиция на $R \iff A_0 = \bigcup_{i=1}^n A_i$, като кортежите в текущия екземпляр на R биват проектирани поатрибутно спрямо схемите на R_i .

Декомпозирането ни дава възможност да се справим с аномалиите и излишествата.

това става по следната стратегия за декомпозиция:

Стратегия за декомпозиция

Търсим нетривиална ФЗ $A \rightarrow B$, т.че:

- $A := \{A_1, \dots, A_n\}$
- $B := \{B_1, \dots, B_m\}$
- A не е суперключ

Тогава декомпозираме $R(A, B, C)$ на релациите S, T със схеми $S(C, A)$ и $T(A, B)$

забл. М/у A, B, C може да има сечения.

def Казваме, че декомпозицията на R на релациите S и T е без загуба тогава и само тогава, когато $S \bowtie T = R$.

Нормални форми

Всяка следваща НФ наследява предишната.

Първа NF

- единственото изискване е атомарност на атрибутите на релацията, което при релационния модел е тривиално изпълнено

Втора NF

def Релацията R със схема $R(A)$ е в 2NF \iff всеки неключов атрибут е в пълна функционална зависимост от който и да е кандидат ключ.

Трета NF

def Релацията R със схема $R(A)$ е в 3NF \iff за всяка нетривиална ФЗ $X \rightarrow Y$ е в сила, че X е суперключ или Y е част от ключ.

Бойс-Код NF(BCNF)

def Релацията R със схема $R(A)$ е в BCNF \iff за всяка нетривиална ФЗ $X \rightarrow Y$ е в сила, че X е суперключ.

напр., всяка релация с 2 атрибута е в BCNF, което е просто следствие от вида на възможните ФЗ в такава релация

Многозначни зависимости

Казваме, че X многозначно определя Y и пишем $X \twoheadrightarrow Y \iff$ ако 2 кортежа в произволен екземпляр на една релация съвпадат по всички атрибути на X , то съответстващите им компоненти от множеството атрибути Y могат да бъдат разменени и така получените 2 нови кортежа отново ще бъдат принадлежат на екземпляра.

С други думи:

def
Нека r е произволен екземпляр на релацията със схема $R(A)$
Нека $\alpha \subseteq A$ и $\beta \subseteq A$
 $\alpha \twoheadrightarrow \beta \iff$
 $(\forall t_1, t_2 \in r(R))(\exists t_3, t_4 \in r(R))[t_1[\alpha] = t_2[\alpha] \implies$
 $\quad t_1[\alpha] = t_2[\alpha] = t_3[\alpha] = t_4[\alpha],$
 $\quad t_1[\beta] = t_3[\beta],$
 $\quad t_2[\beta] = t_4[\beta],$
 $\quad t_1[A \setminus (\beta \cup \alpha)] = t_4[A \setminus (\beta \cup \alpha)],$
 $\quad t_2[A \setminus (\beta \cup \alpha)] = t_3[A \setminus (\beta \cup \alpha)]]$

Разбира се, тази дефиниция е предопределена.

Необходим е само единия кортеж, другият се подразбира, поради кванторите за всеобщност.

def MVD $X \twoheadrightarrow Y$ за релацията със схема $R(A)$ се нарича **тривиална** $\iff Y \subseteq X$ или $X \cup Y = A$.

def Една MVD наричаме **нетривиална** тогава и само тогава, когато не е тривиална.

! **забл.** Всяка ФЗ е МФЗ.

Аксиоми/Правила за многозначните зависимости

Нека имаме релацията със схемата $R(A)$, където A е множеството от атрибутите ѝ.

Транзитивност

Ако е в сила $X \rightarrow\rightarrow Y$ и $Y \rightarrow\rightarrow Z$, където $X, Y, Z \subseteq A$, то $X \rightarrow\rightarrow Z$.

Допълнение

Ако е в сила $X \rightarrow\rightarrow Y$, където $X, Y \subseteq A$ и $C = A \setminus (X \cup Y)$, то $X \rightarrow\rightarrow C$.

Обединение

Ако е в сила $X \rightarrow\rightarrow Y$ и $X \rightarrow\rightarrow W$, където $X, Y, W \subseteq A$, то $X \rightarrow\rightarrow YW$.

! забл. не можем да делим нито лявата, нито дясната страна на дадена МФЗ и да сме сигурни, че резултатните МФЗ са коректни.

! забл. стратегията за декомпозиция остава същата, както тази при функционалните зависимости

Четвърта нормална форма.

def Релацията със схема $R(A)$, където A представлява мв-вото от атрибутите й, е в 4NF \iff за всяка нетривиална функционална зависимост $X \rightarrow\rightarrow Y$ е в сила, че X е суперключ.

Съединение без загуба и случай на загуба

- Прилагайки рекурсивно стратегията за декомпозиция до получаване на декомпозиция на изходната релация, релациите в която удовлетворяват желаната нормална форма, можем да загубим някои от ФЗ. Декомпозирането до 3NF включително е гарантирано без загуба, но при по-строгите NF има случаи, при които резултатната декомпозиция има загуба.

пример за декомпозиция на релация и съединение без загуба

Нека имаме релацията **StudentsInfo(ФН, Име, Град, Област, Премет, Оценка)** и един неин екземпляр:

ФН	Име	Град	Област	Предмет	Оценка
82201	Силвия Х.	Стара Загора	Стара Загора	ДААП	6
82165	Радослав Хърлев	Перник	Перник	ЛП	6
82154	Георги Хърлев	Ловеч	Ловеч	БД	5.5
88888	Мартин Попов	Добрич	Добрич	КАРХ	3
88888	Мартин Попов	Добрич	Добрич	ДААП	5

В тази релация очевидно има много излишества, които потенциално биха довели до аномалии.

Нетривиални ФЗ, които са в сила за релацията:

- ФН, Предмет \rightarrow Име, Град, Област, Оценка (следователно **ФН, Предмет** представлява суперключ; освен това е минимален, следователно е ключ)
- ФН \rightarrow Име, Град, Област
- Град \rightarrow Област

Релацията е в 1НФ, т.к. всеки атрибут има атомарно значение, като заб., че имената, макар и 3 на бр. за всеки атрибут **Име** на кортеж, се пазят като символен низ.

Релацията обаче не е в 2НФ, т.к. не всеки атрибут е в пълна функционална зависимост от ключа.

Напр. ФЗ **2. ФН** → **Име, Град, Област**

Да декомпозираме до **R1(ФН, Име, Град, Област)** и **R2(ФН, Предмет, Оценка)**

- R1(ФН, Име, Град, Област)

ФН	Име	Град	Област
82201	Силвия Х.	Стара Загора	Стара Загора
82165	Радослав Хърлев	Перник	Перник
82154	Георги Хърлев	Ловеч	Ловеч
88888	Мартин Попов	Добрич	Добрич

Тук ФЗ:

1.1. ФН → Име, Град, Област

1.2. Град → Област

Релацията R1 не е в 2NF, т.к. не всеки атрибут е в пълна функционална зависимост от ключа, както се вижда от **1.2**.

- R2(ФН, Предмет, Оценка) ✓

ФН	Предмет	Оценка
82201	ДААП	6
82165	ЛП	6
82154	БД	5.5
88888	КАРХ	3
88888	ДААП	5

Тук ФЗ:

2.1. ФН, Предмет → Оценка

Релацията R1 е в 2NF, т.к. всеки атрибут е в пълна функционална зависимост от ключа.

Релацията е и 3NF, т.е. за всяка нетривиална ФЗ е в сила, че лявата страна е суперключ или дясната е част от ключ, дори в частност е в сила че за всяка ФЗ лявата страна е суперключ, следователно релацията е в BCNF.

Да декомпозираме R1(ФН, Име, Град, Област) на **R3(Град, Област)** и **R4(ФН, Име, Град)**

- R3(Град, Област) ✓

Град	Област
Стара Загора	Стара Загора
Перник	Перник
Ловеч	Ловеч
Добрич	Добрич

Тук ФЗ:

3.1. Град → Област

Релацията е в BCNF, понеже всяка релация с два атрибута е в BCNF.

- R4(ФН, Име, Град) ✓

ФН	Име	Град
82201	Силвия Х.	Стара Загора
82165	Радослав Хърлев	Перник
82154	Георги Хърлев	Ловеч
88888	Мартин Попов	Добрич

Тук ФЗ:

4.1. ФН \rightarrow Име, Град

Релацията е в BCNF, понеже лявата страна на всяка нетривиална ФЗ е суперключ.

Окончателно получихме от **StudentsInfo(ФН, Име, Град, Област, Премет, Оценка)**:

- R2(**ФН**, **Предмет**, Оценка)
- R3(**Град**, Област)
- R4(**ФН**, Име, Град)

и е в сила, че $R2 \bowtie (R4 \bowtie R3) = StudentsInfo$, т.е. тук съединението е без загуба; функционалните зависимости се запазват

пример за декомпозиция на релация и съединение със загуба

Нека имаме релацията с релационна схема $Booking(title, theater, city)$, като имаме следните ФЗ:

1. theater \rightarrow city
2. title, city \rightarrow theater

Така кандидат-ключовете са theater, {city, title} и {theater, title}.

Релацията не е в BCNF, т.к. **1. theater \rightarrow city** е ФЗ, лявата страна на която не е суперключ.

Да декомпозираме на R1(**theater**, city) и R2(**theater**, **title**).

- ФЗ в R1(**theater**, city)
 - 1.1. theater \rightarrow city
- ФЗ в R2(**theater**, city)
 - всички са тривиални

Загубихме ФЗ **2. title, city \rightarrow theater**

Така това ограничение вече не е наложено по никакъв начин и можем да добавим следните кортежи в таблиците:

- R1

Theater	City
El Capitan	LA
New Bev. Cinema	LA

- R2

Theater	Title
El Capitan	The Net
New Bev. Cinema	The Net

При естествено съединение получаваме релация, чиято релационна схема е като на изходната, но имаме нарушение точно на ФЗ, която изгубихме:

Theater	City	Title
El Capitan	LA	The Net
New Bev. Cinema	LA	The Net