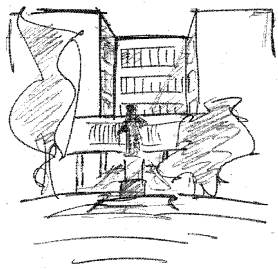


[P273]
Пројектовање база података

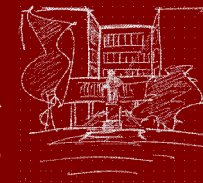
8



Саша Малков
Универзитет у Београду
Математички факултет
2023/2024

[P273]
Пројектовање база података

Саша Малков



Тема 8.5

Логичко моделирање
-
Основне нормалне форме

[P273] Пројектовање база података - Саша Малков - 2023/24 - час 8

1

Логичко моделирање / Нормалне форме

Нормалне форме



- *Нормалне форме* су специфични облици релација који задовољавају одређена правила у вези функционалних зависности
 - који зато *гарантују* да у релацији неће бити редундантности одређеног типа

Универзитет у Београду - Математички факултет

[P273] Пројектовање база података - Саша Малков - 2023/24 - час 8

2

Логичко моделирање / Нормалне форме

Нормалне форме



- У свом првом раду о релационом моделу (1970), Код је дефинисао само
 - 1. нормалну форму (*1NF*)
- Годину дана касније (1971) је дефинисао
 - 2. нормалну форму (*2NF*)
 - 3. нормалну форму (*3NF*)
- Заједно са Рејмондом Бојсом је дефинисао (1974)
 - Бојс-Кодову нормалну форму (*BCNF*)
- Касније су формулисане још и
 - 4. нормална форма (*4NF*) (1977)
 - 5. нормална форма (*5NF*) (1979)
 - Нормална форма домена и кључа (*DKNF*) (1981)
 - Нормална форма елементарног кључа (*EKNF*) (1982)
 - 6. нормална форма (*6NF*) (2003)
 - Нормална форма есенцијалних торки (*ETNF*) (2012)
 - Нормална форма наткључева (*SKNF*)
 - Нормална форма без редунданси (*RNFN*)

Универзитет у Београду - Математички факултет

[P273] Пројектовање база података - Саша Малков - 2023/24 - час 8

3



Нормалне форме (2)

- Већина нормалних форми (све осим $6NF$) могу да се поређају тако да наредна подразумева све претходне:
 - 1. нормална форма ($1NF$)
 - 2. нормална форма ($2NF$)
 - 3. нормална форма ($3NF$)
 - Нормална форма елементарног кључа ($EKNF$)
 - Бојс-Кодова нормална форма ($BCNF$)
 - 4. нормална форма ($4NF$)
 - Нормална форма есенцијалних торки ($ETNF$)
 - Нормална форма без редунданси ($RFNF$)
 - Нормална форма суперкључева ($SKNF$)
 - 5. нормална форма ($5NF$)
 - Нормална форма домена и кључа ($DKNF$)
- 6. нормална форма ($6NF$) представља алтернативу за $DKNF$ и не може да се непосредно упореди са осталим нормалним формама



1. нормална форма

- Релација је у првој нормалној форми ако сваки атрибут може да има само атомичне вредности.
 - Релациони модел претпоставља атомичне вредности
 - ЕР омогућава и сложене вредности атрибута



1. нормална форма (пример)

- Нека један предмет може да држи више наставника
- *Неатомични* модел би омогућавао да у релацији *предмет* постоји атрибут *наставници* са именима више наставника
- $1NF$ то забрањује
- Једно могуће решење је да се атрибут *наставници* замени већим бројем атрибута *наставник1*, *наставник2*, *наставник3*
 - такво решење није добро ни ако има мање ни ако има више од 3 наставника за неки предмет, па чак ни када их је тачно 3
- Боље решење је да се уведе нова релација која моделира однос предмета и наставника



2. нормална форма

- Релација је у другој нормалној форми
 - ако је у првој нормалној форми и
 - ако ниједан некључан атрибут (тј. који не припада ниједном кандидату кључу) није функционално зависан од неког правог подскупа атрибута неког кандидат кључа.
- Суштина је да у случају сложеног кључа сви остали атрибути зависе од *цело* кључа а не од неког његовог дела.
 - т.ј. кандидати кључеви су минимални



2. нормална форма (пример)

- Нека следећа релација описује предмете које су уписали студенти:
 - УписанПредмет{ **индекс**, **шифр**ред, име, презиме, називред, којијуи }
 - Релација је у 1.НФ
- Релација није у 2.НФ
 - *име* и *презиме* зависе од атрибута *индекс*
 - *називред* зависи од атрибута *шифр*ред
- Да би се задовољила 2.НФ потребно је да се релација подели на више релација



2. нормална форма (пример)

- Релацију
 - УписанПредмет{ **индекс**, **шифр**ред, име, презиме, називред, којијуи }
- Можемо да поделимо на 3 релације:
 - УписанПредмет{ **индекс**, **шифр**ред, којијуи }
 - Студент{ **индекс**, име, презиме }
 - Предмет{ **шифр**ред, називред }
- Према теорему о декомпозицији, декомпозиција је потпуна
 - Показује се за једну по једну декомпозицију, најпре:
 - УписанПредмет{ **индекс**, **шифр**ред, називред, којијуи }
 - Студент{ **индекс**, име, презиме }
 - а затим и
 - УписанПредмет{ **индекс**, **шифр**ред, којијуи }
 - Предмет{ **шифр**ред, називред }



2. нормална форма (дискусија)

- Нарушавање 2.НФ указује на неисправно груписање атрибута у релације
 - релација моделира више различитих ФЗ
 - то води редувантности
- Постизање 2.НФ је кључно за позиционирање атрибута у *одговарајуће* релације
- Ипак, 2.НФ допушта да у релацији постоје два кључа кандидата, као и транзитивне зависности
- Увек је могућа потпуна декомпозиција до 2НФ



3. нормална форма

- Ако је R релација и X неки подскуп њених атрибута и A неки атрибут релације R , онда је R у 3.НФ ако за сваку ФЗ $X \rightarrow A$ на релацији R важи једно од:
 - $A \in X$, тј. A припада скупу X (тзв. тривијална зависност)
 - X је наткључ (тј. кључ или садржи кључ)
 - A је део неког кључа релације
- Оригинална дефиниција (Код, 1971)
 - Релација је у 3.НФ ако је
 - у 2.НФ и
 - сваки неключан атрибут је нетранзитивно зависан од сваког кључа.



3. нормална форма (2)

- У основи налаже да сваки неключни атрибут може да представља само додатни опис неког кључа
- Тј. нема транзитивних зависности међу неключним атрибутима (неключни атрибути не представљају описе неключних атрибута)
- Могућа нарушавања 3.НФ:
 - X је подскуп неког кључа K
 - тј. имамо зависност од дела кључа
 - нарушена је и 2.НФ
 - X није ни подскуп ни надскуп ниједног кључа
 - тј. имамо транзитивну зависност



3. нормална форма (пример)

- Нека следећа релација описује уписане студенте:
 - $Студент\{ \underline{индекс}, име, презиме, студијско\text{ број}, ниво, \text{ испрајање} \}$
- Релација јесте у 2.НФ али није у 3.НФ
 - $ниво$ и $испрајање$ зависе од неключног атрибута $студијско\text{ број}$
- Да би се задовољила 3.НФ потребно је да се релација подели на 2 релације:
 - $Студент\{ \underline{индекс}, име, презиме, студијско\text{ број} \}$
 - $Програм\{ \underline{студијско\text{ број}}, ниво, \text{ испрајање} \}$



3. нормална форма (дискусија)

- Нарушавања 3.НФ указује да је у релацији моделирано више релативно независних ФЗ
- Увек је могућа потпуна декомпозиција до 3.НФ



Нормална форма елементарног кључа

- Релација је у НФЕК ако је у 3.НФ и ако за сваку елементарну ФЗ $X \rightarrow Y$ на релацији R важи једно од:
 - X је кључ или
 - Y је елементарни кључ или део елементарног кључа
- При томе:
 - ФЗ $X \rightarrow Y$ је *елементарна* ако не постоји ФЗ $Z \rightarrow Y$, ни за који прави подскуп атрибута $Z \subset X$
 - Кључ је *елементаран* ако је домен бар једне елементарне функционалне зависности
- Ова форма стоји између 3.НФ на БКНФ
- Пример и мотивација ће бити објашњени након БКНФ



Бојс-Кодова нормална форма

- Ако је R релација и X неки подскуп њених атрибута и A неки атрибут релације R , онда је R у Бојс-Кодовој нормалној форми ако за сваку ФЗ $X \rightarrow A$ на релацији R важи једно од:
 - $A \in X$, тј. A припада скупу X , тзв. тривијална зависност
 - X је наткључ
- Тј.
 - сваки атрибут релације је или део кључа или зависи од *цело*т кључа
 - не постоје зависности међу некључним атрибутима ни транзитивне зависности



Бојс-Кодова нормална форма, дискусија

- Релације које су у З.НФ најчешће задовољавају и БКНФ
 - Практично увек када постоји само један кључ или нема преклапања међу кључевима
 - Већина примера “релација у ЗНФ које нису у БКНФ”, који се могу пронаћи у литератури, заправо нису исправни, тј. обично нису у ЗНФ иако аутори тврде супротно
- Штавише, тешко је пронаћи добар пример који
 - јесте у З.НФ
 - није у БКНФ
 - може да се преведе у БКНФ



БКНФ, пример 1.1

- (https://en.wikipedia.org/wiki/Boyce-Codd_normal_form)
- Посматрамо релацију
 - Резервација { шерен, о~~г~~, го, шарифа }
 - постоје два терена, један је травнати и један бетонски
 - резервације су дефинисане тереном и периодом
 - постоје 4 тарифе – по две за сваки од терена зависно да ли је термин предвиђен за чланове или не
- Релација јесте у 2.НФ и 3.НФ (штавише, нема атрибута који су ван кључева)
 - постоји већи број кључева
 - { шерен, о~~г~~, го, шарифа }
 - { шерен, о~~г~~, го, шарифа }
 - { шерен, о~~г~~, го, шарифа }
 - { шерен, о~~г~~, го, шарифа }
- Релација *није* у БКНФ
 - проблем је у зависности шарифа \rightarrow шерен
 - шарифа није наткључ
 - шерен не припада скупу { шарифа }



БКНФ, пример 1.2

- Једно решење је додавање новог атрибута за_чланове:
 - Резервација { шерен, о~~г~~, го, шарифа, за_чланове }
- Сада постоје додатне ФЗ:
 - { шарифа } \rightarrow { шерен, за_чланове }
 - { шерен, за_чланове } \rightarrow { шарифа }
- Користимо другу за потпуну декомпозицију:
 - Резервација { шерен, о~~г~~, го, за_чланове }
 - Тарифа { шерен, за_чланове, шарифа }
 - може и обрнуто, да кључ буде шарифа, зато што је ФЗ у оба смера
- Испоставља се да је проблем у *скривеним* ФЗ које постају очигледне када се додају нови атрибути



БКНФ, пример 2.1

- Посматрамо релацију
 - $R_{\text{својоред}}\{ \text{радник}, \text{тим}, \text{руководилац} \}$
- и зависности
 - $\text{радник}, \text{тим} \rightarrow \text{руководилац}$
 - сваки радник може да буде у више тимова
 - у сваком тиму радник има тачно једног руководиоца
 - у једном тиму може да буде више руководилаца
 - али су одговарајући скупови подређених радника дисјунктни
 - $\text{руководилац} \rightarrow \text{тим}$
 - руководилац може да руководи само у једном тиму



БКНФ, пример 2.2

- (Teorey, 2011) наводи овај пример:
 - “релација јесте у 3.НФ а није у БКНФ”
 - проблем је у аномалији брисања – ако сви радници код неког руководиоца напусте тим, онда ће се изгубити информација о руководиоцу у тиму
 - предлаже да се проблем реши тако да се додатно уведе релација $R_{\text{руководилац}}\{ \text{руководилац}, \text{тим} \}$ која би описивала другу зависност



БКНФ, пример 2.3

- Претходни проблем се делимично решава ако учимо да имамо још једну зависност:
 - $\text{радник}, \text{руководилац} \rightarrow \text{тим}$
 - ова зависност не изгледа *природно*, тим пре што може да се изведе из постојећих, али је она заправо непосредна последица
 - ограничења да сваки руководилац руководи највише једним тимом
 - и претпоставке да сваки радник има руководиоца у сваком тиму у који је распоређен
- Ако ту Ф3 прогласимо за основну, а раније установљену $\text{радник}, \text{тим} \rightarrow \text{руководилац}$ за изведену...
- ...онда видимо да није задовољена чак ни 2.НФ (а самим тим ни 3.НФ):
 - $\text{радник}, \text{руководилац} \rightarrow \text{руководилац} \rightarrow \text{тим}$



БКНФ, пример 2.4

- Да бисмо задовољили 2.НФ морамо да направимо две релације:
 - $R_{\text{својоред}}\{ \text{радник}, \text{руководилац} \}$
 - $R_{\text{руководилац}}\{ \text{руководилац}, \text{тим} \}$
- Сада су задовољене и 2.НФ и 3.НФ и БКНФ
- Решење је једноставније од изворног
 - у релацији $R_{\text{својоред}}$ нема атрибута *тим*
- ...али је изгубљено ограничење да у сваком тиму радник има тачно једног руководиоца
 - ...зато што смо напустили првобитну Ф3
- Видимо да полазни пример није задовољавао 2.НФ (а тиме ни 3.НФ) зато што нису довољно пажљиво разматране функционалне зависности
- Али видимо и да када додајемо нове зависности не смемо да занемаримо старе, иначе можемо да их изгубимо



БКНФ и НФЕК

- Нека имамо релацију:
 - ПријаваИспитиџа (стџудентџ, исџиџни рок, џредметџ)
 - кључ је { стџудентџ, исџиџни рок }
- и зависности:
 - { стџудентџ, исџиџни рок } → { џредметџ }
 - тј. студент у једном исп. року може да пријави испит из највише једног предмета
 - { џредметџ } → { исџиџни рок }
 - тј. сваки предмет се полаже само у по једном испитном року
- Тада:
 - релација ПријаваИспитиџа није у БКНФ:
 - испитни рок зависи од предмета, а предмет није део кључа
 - ако направимо декомпозицију, нарушићемо прву зависност...
 - ако променимо кључ (стџудентџ, џредметџ), проблем опстаје



БКНФ и НФЕК (2)

- Али:
 - релација ПријаваИспитиџа **јест** у НФЕК
 - прва зависност је од целог кључа
 - друга је према делу кључа
 - обе су елементарне



БКНФ и НФЕК (3)

- Проблем не може добро да се реши као у претходном примеру, увођењем алтернативних зависности
- Постоји потпуна декомпозиција:
 - ПријаваИспитиџа (стџудентџ, џредметџ)
 - ПредметџУРоку (џредметџ, исџиџни рок)
- Она је потпуна, али **не чува све зависности**
 - није очувано ограничење да у једном испитном року студент може да пријави испит из највише једног предмета
- Штавише, не постоји декомпозиција која чува све ФЗ



БКНФ и НФЕК (4)

- Значај НФЕК је у томе што неке релације не могу да пређу пут од ЗНФ до БКНФ а да се очувају све зависности...
- ...али могу да дођу бар до НФЕК
- Да би се у моделу сачувале све зависности, потребно је да се постојећој релацији дода нова релација која описује другу зависност:
 - ПредметџУРоку (џредметџ, исџиџни рок)
 - Уводи се редундантност...
 - ...која се решава додавањем страног кључа (предмет, испитни рок) из релације ПријаваИспитиџа према релацији ПредметџУРоку
 - (што не подржавају сви РСУБП, зато што се страни кључ не везује за примарни)



Пречишћавање схеме

- На претходном часу смо обрадили уводне опште делове ове теме, као и:
 - функционлне зависности
 - основне нормалне форме
- Данас ћемо да обрадимо преостале теме из ове области:
 - **вишезначне зависности**
 - **више нормалне форме**
 - **поступак нормализације**

Пројектовање база података

Саша Малков



Тема 8.6

Логичко моделирање

-

Вишезначне зависности и зависности спајања



Ограничења функционалних зависности

- Као што њихово име каже, ФЗ описују зависности које имају облик функције (пресликавања):
 - за дату вредност једног (скупа) атрибута одређују тачно једну вредност другог (скупа) атрибута
- То није довољно за описивање свих зависности у неком посматраном домену или у БП која га моделира



Ограничења функционалних зависности (2)

- На пример:
 - интуитивно је јасно да би требало да постоји нешто као ФЗ између индекса студента и скупа уписаних школских година
 - али на десној страни је *скуп*, а не тачно једна вредност
 - то не може да се опише помоћу ФЗ
- Или, на пример:
 - ако посматрамо, са једне стране, студента и школску годину, а са друге стране скуп предмета које је студент уписао у једној школској години
 - све је исто као у претходном примеру, осим што домен зависности није један атрибут (индекс студента) већ пар који чине индекс студента и школска година



Ограничења функционалних зависности (3)

- Формална потврда да ФЗ нису довољне је чињеница да:
 - ФЗ јесу довољан услов за потпуну декомпозицију
 - ФЗ *нису* неопходан услов за потпуну декомпозицију
- На пример, нека релација илуструје податке о погледаним филмовима:

Гледалац	Наслов	Глумац
Горан	Балкански шпијун	Данило Бата Стојковић
Драгана	Балкански шпијун	Данило Бата Стојковић
Горан	Балкански шпијун	Бора Тодоровић
Драгана	Балкански шпијун	Бора Тодоровић
Горан	Професионалац	Бора Тодоровић
Горан	Професионалац	Бранислав Лечић

- Не постоји ФЗ, али је могућа потпуна декомпозиција на:
 - {Гледалац, Наслов}, {Наслов, Глумац}



Вишезначне зависности

- Вишезначне зависности** се уводе зато што ФЗ нису довољне да опишу све редундантности
 - а тиме ни све услове за потпуну декомпозицију
- Идеја је да свака вредност једног скупа атрибута X одређује **скуп вредности** другог скупа атрибута Y
 - функционалне зависности представљају специјалан случај вишезначних зависности, када је сваки зависан скуп једночлан



Вишезначне зависности (деф)

- Дефиниција
 - Између скупа атрибута X и зависног скупа атрибута Y релације R (где су $X, Y \subseteq Attr(R)$ и $Z = Attr(R) \setminus XY$) постоји **вишезначна зависност**, у ознаци $X \twoheadrightarrow Y$, ако за сваке две торке t_1 и t_2 из R такве да је $t_1[X] = t_2[X]$ постоји торка t_3 таква да је $t_3[X] = t_1[X] = t_2[X]$ и $t_3[Y] = t_1[Y]$ и $t_3[Z] = t_2[Z]$.
 - користићемо и ознаку $X \twoheadrightarrow Y$



Вишезначне зависности (2)

- Неформално речено, за сваку изабрану вредност атрибута из X , вредности из Y су независне од вредности из Z , тј. зависе само од X .
- Ако постоје торке (x, y_1, z_1) и (x, y_2, z_2) , онда постоји и торка (x, y_1, z_2)
 - последица је да постоји и торка (x, y_2, z_1)
- Т.ј. ако су X, Y и Z дисјунктни скупови атрибута који покривају $Attr(R)$, онда важи $\forall X \twoheadrightarrow Y$ ако за сваку фиксирану вредност x рестрикција по услову $X=x$ && $Z=z$ има исту вредност, без обзира на изабрано z .



Вишезначне зависности (пример)

- На претходном примеру можемо да уочимо вишезначне зависности:
 - $\{\text{Наслов}\} \twoheadrightarrow \{\text{Глумац}\}$
 - $\{\text{Наслов}\} \twoheadrightarrow \{\text{Гледалац}\}$

Гледалац	Наслов	Глумац
Горан	Балкански шпијун	Данило Бата Стојковић
Драгана	Балкански шпијун	Данило Бата Стојковић
Горан	Балкански шпијун	Бора Тодоровић
Драгана	Балкански шпијун	Бора Тодоровић
Горан	Професионалац	Бора Тодоровић
Горан	Професионалац	Бранислав Лечић



Особине вишезначних зависности

- Ако су X, Y, W и Z скупови атрибута на R такви да је $W = Attr(R) \setminus XY\dots$
- **Комплементарност**
 - Ако важи $X \twoheadrightarrow Y$ онда важи и $X \twoheadrightarrow W$
- **Рефлексивност**
 - Ако важи $Y \subseteq X$, онда важи и $X \twoheadrightarrow Y$
- **Проширивост**
 - Ако важи $X \twoheadrightarrow Y$ онда важи и $XZ \twoheadrightarrow YZ$
- **Транзитивност**
 - Ако на R важи $X \twoheadrightarrow Y$ и $Y \twoheadrightarrow Z$ онда важи и $X \twoheadrightarrow Z \setminus Y$



Веза вишезначних зависности и ФЗ

- **Конвертибилност**
 - Ако на R важи $X \twoheadrightarrow Y$ онда важи и $X \twoheadrightarrow Y$
 - Т.ј. ФЗ су спец. случај ВЗ
- **Инијеративност**
 - Ако на R важи $X \twoheadrightarrow Y$ и $XY \twoheadrightarrow Z$, онда важи и $X \twoheadrightarrow Z \setminus Y$
 - претпоставка је да су сви скупови атр. непразни



Потпуност наведених особина

- **Теорема:**
 - Ако је D скуп ВЗ (а тиме и ФЗ) на R , онда се скуп D^+ свих ВЗ логички изводивих из D може извести и применом само наведених особина ВЗ и аксиома ФЗ



ВЗ и декомпозиција

- **Теорема:**
 - Нека су X, Y, Z непразни скупови атрибута релације R , и при томе је $Z = \text{Attr}(R) \setminus XY$.
Онда на релацији R важи једнакост $R = R[XY] * R[XZ]$ акко важи ВЗ $X \rightarrow Y$
- Еквивалентно тврђење:
 - Релација R се може потпуно декомпоновати на $R[W]$ и $R[Q]$ акко важи ВЗ: $W \cap Q \rightarrow W \setminus Q$ (акко важи ВЗ: $W \cap Q \rightarrow Q \setminus W$)



ВЗ и декомпозиција (2)

- Веома је важно што постоји еквиваленција (акко) између ВЗ и потпуне декомпозиције једне релације на две релације
- Због тога је установљавање ВЗ довољан и неопходан услов за потпуну декомпозицију једне релације на две мање релације
- Приметимо “две”...



Тривијалне ВЗ

- Ако је X непразан скуп атрибута, онда кажемо да су ВЗ $X \rightarrow \emptyset$ и $X \rightarrow \text{Attr}(R) \setminus X$ **тривијалне**



ВЗ нису довољне

- У неким случајевима једна релација не може да се потпуно декомпонује на две релације али може на три
 - тј. ВЗ ипак не представљају најопштији облик зависности



Зависности спајања

- Нека су X_1, X_2, \dots, X_n скупови атрибута релације R , чија је унија управо $Attr(R)$. У релацији R важи **зависности спајања**, у ознаци $\{X_1, X_2, \dots, X_n\}$ ако је:

$$R = R[X_1] * R[X_2] * \dots * R[X_n]$$
- Зависности спајања су дефинисане тако да представљају најопштији могући облик зависности – неопходан и довољан за потпуну декомпозицију
 - зависности спајања са два скупа атрибута су еквивалентне вишезначним зависностима



Тривијална зависност спајања

- Зависност спајања $\{X_1, X_2, \dots, X_n\}$ је тривијална ако за неки од скупова атрибута X_i важи $X_i = Attr(R)$.



Редуцибилна зависност спајања

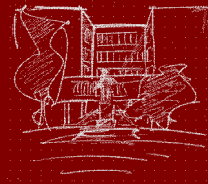
- Зависност спајања је редуцибилна ако постоји прави подскуп њених компоненти, који и даље одређује (редуковану) зависност спајања
 - еквивалентно:
 - Зависност спајања $\{X_1, X_2, \dots, X_n\}$ је редуцибилна ако нека компонента X_i може да се искључи, а да и даље имамо зависност спајања $\{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}$.
- На пример, ако је X_i прави подскуп неке друге компоненте, онда је ЗС редуцибилна



Редуцибилна зависност спајања (2)

- Зависност спајања је нередуцибилна ако није редуцибилна
- Занимају нас само нередуцибилне зависности спајања
 - Свака редуцибилна ЗС може да се сведе на комбинацију нередуцибилне ЗС и тривијалних ЗС

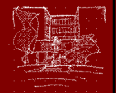
[P273]
Пројектовање база података
Саша Малков



Тема 8.7

Логичко моделирање
-
Више нормалне форме

Више нормалне форме



- Првих неколико нормалних форми смо дефинисали на основу функционалних зависности:
 - 1. НФ
 - чак и без функционалних зависности
 - 2. НФ
 - 3. НФ
 - НФЕК
 - БКНФ

Више нормалне форме (2)



- Остале НФ се називају и **више нормалне форме**
 - Дефинишу се на основу сложенијих облика зависности
- Већина је уређена по сложености (виша имплицира нижу):
 - 4.НФ
 - НФЕТ – нормална форма есенцијалних торки (ETNF)
 - НФБР – нормална форма без редунданси (RFNF)
 - НФНК – нормална форма наткључева (SKNF)
 - 5.НФ
 - НФДК – нормална форма домена и кључа (DKNF)
- 6.НФ не може да се “поређа”

4. нормална форма



- Релација R је у 4.НФ ако за сваку вишезначну зависност $X \rightarrow Y$ важи да је X кључ или наткључ.
- У суштини, ако релација не задовољава 4.НФ, то значи да се може дефинисати као резултат спајања двеју једноставнијих релација



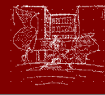
4. нормална форма (пример 1)

- Нека имамо релацију из ранијег примера
 - *Филм*(гледалац, наслов, глумац)
- Видели смо да важе зависности
 - $\{ \text{наслов} \} \twoheadrightarrow \{ \text{гледалац} \}$
 - $\{ \text{наслов} \} \twoheadrightarrow \{ \text{глумац} \}$
 - нарушена је 4НФ, зато што $\{ \text{наслов} \}$ није (нат)кључ
- онда се *Филм* може нормализовати декомпозицијом на:
 - *ФилмГледалац*(наслов, гледалац)
 - *ФилмГлумац*(наслов, глумац)



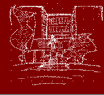
4. нормална форма (пример 2)

- Нека имамо релацију
 - *Понуда*(продавница, производ, град)
- и нека важи правило
 - $\{ \text{продавница} \} \twoheadrightarrow \{ \text{производ} \}$
 - т.ј. исти скуп производа се нуди независно од града
 - т.ј. у сваком граду се нуде сви производи које продавница нуди
 - наравно, онда мора да важи и $\{ \text{продавница} \} \twoheadrightarrow \{ \text{град} \}$
 - нарушена је 4НФ, зато што $\{ \text{продавница} \}$ није (нат)кључ
- онда се *Понуда* може нормализовати декомпозицијом на:
 - *ПонудаПроизвод*(продавница, производ)
 - *ПонудаГрад*(продавница, град)



4. нормална форма (дискусија)

- Евентуална незадовољеност 4.НФ је релативно лако интуитивно уочљива пажљивим посматрањем
 - зато се често заобилази у основној настави база података
 - као и вишезначне зависности
 - очекује се да ће проблем бити уочен и решен интуитивно
 - али то ипак није случај и овакви проблеми се појављују у продукционим базама података
- Незадовољавање 4.НФ може да доведе до драстичног нарушавања перформанси, због значајног повећавања броја редова у конкретним табелама...
- ...наравно, и до проблема са одржавањем услед редундантности



Нормална форма есенцијалних торки

- Релација R задовољава НФЕТ ако свака торка ове релације мора да буде *есенцијална*.
 - Торка је *есенцијална* ако није ни потпуно ни делимично редундантна.
 - Торка t је потпуно редундантна у релацији R ако постоји неки скуп торки S релације R ($t \notin S$) такав да се применом ФЗ и правила интегритета на тај скуп може логички доказати да торка t мора да постоји у релацији R .
 - Торка t је делимично редундантна у релацији R ако у истој релацији постоји ФЗ $X \rightarrow A$ и торка q ($t \neq q$) таква да су им пројекције на X једнаке: $t[X] = q[X]$.
 - т.ј. постоје две различите торке у којима се ФЗ примењује на исти начин, на исте вредности X и A
- Важи (теорема):
 - Нека је релација R одређена само функционалним зависностима и зависностима спајања. Релација R је у НФЕТ ако је у БКНФ и свака зависност спајања има неку компоненту која је наткључ.
- (Darween et al., 2012)



Нормална форма есенцијалних торки (2)

- Неформално, ако посматрамо релацију као скуп чињеница
 - торка је есенцијална ако носи суштински нову информацију, која без ње не може да се израчуна
 - торка је потпуно редундантна ако цела може да се израчуна из других торки
 - торка је делимично редундантна ако неки њен део може да се израчуна из других торки и њеног преосталог дела



НФЕТ – пример (Darween et al., 2012)

- Посматрамо релацију $R \{ S, P, J \}$
 - S – идентификатор снабдевача
 - P – идентификатор производа
 - J – идентификатор пројекта
 - “снабдевач S испоручује производ P за пројекат J ”
 - где постоји зависност спајања $\{ SP, PJ, JS \}$
 - и ФЗ: $SP \rightarrow J \equiv$ “снабдевач испоручује неки део за највише један пројекат”
 - она је у 4.НФ
 - нема друге ФЗ
 - има само тривијалне ВЗ
 - ЗС не може да се изведе из ФЗ
 - ова релација јесте у НФЕТ, а није у 5.НФ (видећемо касније)
- нека садржи торке (s, p, j') , (s, p', j) , (s', p, j)
- онда по зависности спајања мора да садржи и (s, p, j)
 - али по ФЗ мора да важи $j=j'$
 - па (s, p, j) није потпуно редундантна торка



НФЕТ – значај

- Основни значај НФЕТ је у томе што је
 - нижа нормална форма од 5.НФ
 - у потпуности елиминира редундантне торке
 - могућа је редундантност на нивоу атрибута, коју превазилази НФБР
- Представља теоријски резултат (са НФБР) којим се показује да за потпуну елиминацију редундантности не мора да се оствари 5.НФ



Нормална форма без редунданси

- Релација R задовољава НФБР ако не може да постоји инстанца релације (садржај) која садржи неку торку t и неки атрибут A тако да је пар (t, A) редундантан.
 - Вредност атрибута A торке $t: (t, A)$ је редундантна у релацији ако за било коју торку q , која се разликује од t само у вредности атрибута A , инстанца релације која би се добила када би се торка t заменила са q не би представљала исправну инстанцу.
 - Другим речима, вредност атрибута неке торке је редундантна ако у тој релацији не може да се замени само та једна вредност а да релација и даље остане исправна
- (Darween et al., 2012)



Нормална форма без редунданси (2)

- Нормална форма без редунданси значи да је садржај релације минималан – ако би се заменила макар једна вредност атрибута, онда садржај више не би био исправан



Нормална форма без редунданси (пример)

- Нека имамо релацију: *УдеоПрограма*(програм, *удео*)
 - и правило интегритета: сума удела је 100%
 - Пример садржаја: { (I, 50%), (M, 40%), (A, 10%)}
- Ова релација јесте у НФЕТ
 - ниједан ред не може да се израчуна на основу осталих
 - ниједан атрибут се не понавља
- Ова релација није у НФБР
 - атрибут *удео* је редундантан у свакој изабраној торци, зато што може да се израчуна на основу правила интегритета
 - тј. ако се неки удео замени неком другом вредношћу, садржај релације постаје неисправан



Нормална форма без редунданси (3)

- НФБР се разликује од свих осталих по томе што често не може да се решава декомпозицијом
 - Уместо тога, примењује се промена семантике атрибута
- У претходном примеру, проблем би се решио ако се не би водили проценти него број студената
 - пример садржаја: { (I, 200), (M, 160), (A, 40)}
 - проценат се рачуна на основу садржаја
 - ниједна вредност није редундантна



Нормална форма наткључева

- Релација *R* задовољава НФНК ако је свака компонента сваке нередуцибилне зависности спајања наткључ.
- Зависност спајања је нередуцибилна ако не постоји подскуп њених компоненти који одређује (редуковану) зависност спајања
- (Darween et al., 2012)
- Пример ћемо да видимо после 5НФ



5. нормална форма

- Релација R је у 5.НФ ако је свака нетривијална зависност спајања логички имплицирана кључевима релације R .
- Зависност спајања је имплицирана кључевима ако свака компонента зависности спајања представља наткључ релације
- Позната је и под именом Нормална форма пројективног спајања (PJNF).
- Разлика у односу на нормалну форму наткључева је што су зависности “логички имплициране кључевима кандидатима”, а не морају да “представљају наткључеве”
 - у раду (Darween et al., 2012) је доказ да је то строго јачи услов



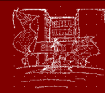
5. нормална форма (пример 1)

- Пример релације у 5.НФ може да буде свака релација у којој су
 - неклучни атрибути (сваки за себе) функционално зависни од кључа
 - нпр. *Студент*(индекс, име, презиме)
 - нетривијална ЗС је $\{ \{ \text{индекс, име} \}, \{ \text{индекс, презиме} \} \}$



5. нормална форма (пример 2)

- Нека поново имамо пример са продавницама:
 - *Понуда*(продавница, производ, град)
- и нека важи правило:
 - ако продавница A испоручује производ B и обавља испоруке у град C , и при томе неко испоручује производ B у град C , онда обавезно и продавница A испоручује производ B у град C
- Релација јесте у 4.НФ – нема вишезначних зависности
 - не постоје $V3 \{ \text{град} \} \twoheadrightarrow \{ \text{производ} \}$ и $\{ \text{град} \} \twoheadrightarrow \{ \text{продавница} \}$ због “ако продавница A испоручује производ B и обавља испоруке у град C ...”
- Релација није у 5.НФ – постоји декомпозиција (зависност спајања):
 - $\{ \{ \text{продавница, производ} \}, \{ \text{продавница, град} \}, \{ \text{производ, град} \} \}$
- Проблем се решава наведеном декомпозицијом на три релације
- Не постоји декомпозиција на две релације
 - иначе би постојала $V3$



5. нормална форма (пример 3)

- Нека имамо пример са продавницама, али са другачијим кључевима:
 - *Понуда*(производ, продавница, град)
 - важе ФЗ:
 - $\{ \text{продавница, град} \} \rightarrow \{ \text{производ} \}$
 - $\{ \text{продавница, производ} \} \rightarrow \{ \text{град} \}$
 - $\{ \text{производ, град} \} \rightarrow \{ \text{продавница} \}$
 - важи ЗС: $\{ \{ \text{производ, град} \}, \{ \text{продавница, град} \}, \{ \text{продавница, производ} \} \}$
- Релација јесте у 4.НФ, тј. све компоненте сваке нередуцибилне ЗС су НК
 - ако би постојала нередуцибилна ЗС чија компонента није НК, била би попут $\{ \{ \text{производ, град} \}, \{ \text{продавница} \} \}$, тј. бар једна компонента мора да буде једночлана и недсадржана у остатку
 - али онда имамо контрадикцију:
 - нека је $R = \{ (\text{Макси, вода, Београд}), (\text{Идеа, хлеб, Ниш}) \}$, тј. два реда и све вредности се разликују
 - по ЗС имамо да би и торке (Макси, хлеб, Ниш) и (Идеа, вода, Београд) припадале R
- Релација није у 5НФ - једини кључеви су домени ФЗ, а они не имплицирају ЗС
 - нека постоје торке (Макси, вода, Београд), (Макси, хлеб, Ниш) и (Рога, вода, Ниш)
 - из ЗС мора да постоји и ред (Макси, вода, Ниш)
 - али то не може да се докаже на основу ФЗ



Нормална форма домена и кључа

- Релација задовољава **НФДК** ако у њој не постоје други услови и ограничења осим услова домена и услова кључа.
- Практично, услови домена прописују допуштене вредности атрибута а услови кључа представљају једине допуштене ФЗ
- Пре него што је Дејт обликовао данас познату 6.НФ, неки аутори су НФДК називали 6.НФ.



6. нормална форма

- Релација је у 6. нормалној форми ако не задовољава ниједну зависност спајања.
 - тј. ако не постоји потпуна декомпозиција
 - то практично значи да је у 5.НФ и садржи примарни кључ и највише још један додатни атрибут који је ФЗ од кључа
- Иако је формално врло чиста, оваква форма је потпуно неприхватљива за трансакционе базе података
 - Ако би релација била у 5. НФ и имала кључ и 20 атрибута, да би се довела до 6. НФ морала би да се подели на 20 релација
 - Уместо једног ажурирања једне релације може да буде потребно да се изврши 20 ажурирања различитих релација



6. нормална форма

- У неким случајевима оваква форма може да има смисла
 - Ако је низак број различитих вредности у колонама онда 6.НФ има смисла за складишта података
 - на том моделу почивају тзв. *колонске* нерелационе базе података
 - на истом концепту почивају и битмапирани индекси
 - ...више о томе на једном од наредних часова...
 - У временским базама података, где се води време промене податка
 - ако се време води за целу торку, онда није јасно на који се атрибут односи
 - ако је релација у 6.НФ, онда је јасно да се време односи или на додатни атрибут, или (ако њега нема) на део примарног кључа без времена
 - (примарни кључ свакако обухвата време)



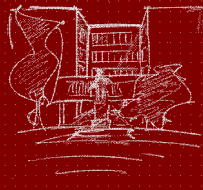
6. нормална форма и НФДК

- 6НФ није сасвим у складу са НФДК
- Проблем је у разбијању на тривијалне релације, јер се губи могућност елементарне провере *пошћуносћи* ентитета
- На пример,
 - нека имамо релацију
 - *Студент* { *индекс*, *име*, *презиме* }
 - и функционалну зависност: *индекс* → { *име*, *презиме* }
 - сваки студент *мора* да има јединствен пар име и презиме
 - релација јесте у НФДК
 - свођењем на 6НФ добијамо две релације
 - { *индекс*, *име* }, { *индекс*, *презиме* }
 - али провера да сваки студент има и име и презиме није елементарна
 - може да се оствари узајамним страним кључевима
 - али то онда више није НФДК

Пројектовање база података

[P273]

Саша Малков



Тема 8.8

Логичко моделирање - Поступци и алгоритми

Нормализација



- Основни циљ пречишћавање схеме у случају релационог модела је елиминација редундантности из схеме базе података
- Средство којим се то остварује је *нормализација логичког модела*
- Нормализација логичког модела базе података је трансформисање логичког модела у еквивалентан скуп релација који задовољава изабране нормалне форме

Нормализација (2)



- Обично се као циљ нормализације поставља БКНФ или 4.НФ
- али може да буде и другачије

Нормализација до БКНФ



- Нормализација логичког модела базе података до БКНФ је свођење на скуп релација које су све у Бојс-Кодовој нормалној форми
- Ефективан алгоритам
 - 1. Нека је R релација која није у БКНФ
 - и нека је X прави подскуп њених атрибута који није наткључ
 - и A један атрибут који зависи од X (и тако нарушава БКНФ), онда је потребно декомпозицијом поделити R на релације са атрибутима XA и $R \setminus A$
 - ако се тако не губе неке ФЗ
 - 2. Ако $R \setminus A$ или XA нису у БКНФ, декомпоновати их даље рекурзивном применом истог алгоритма.



Слабости нормализације до БКНФ

- У неким случајевима **не постоји** декомпозиција у БКНФ која чува све зависности
- тада смо принуђени
 - или да останемо у нижој НФ
 - или да изгубимо неке зависности
 - или да променимо семантику функционалних зависности



Слабости БКНФ – пример

- Нека имамо релацију:
 - **ПријаваИспитића**(*студент*, *испитни рок*, *предмет*)
 - кључ је (*студент*, *испитни рок*)
- и зависности:
 - { *студент*, *испитни рок* } → { *предмет* }
 - тј. студент у једном исп. року може да пријави испит из највише једног предмета
 - { *студент*, *испитни рок* } је кључ
 - { *предмет* } → { *испитни рок* }
 - тј. сваки предмет се полаже само у по једном испитном року
- Тада:
 - релација **ПријаваИспитића** није у БКНФ:
 - испитни рок зависи од предмета, а предмет није део кључа
 - ако направимо декомпозицију, нарушићемо прву зависност...
 - ако променимо кључ (*студент*, *предмет*), проблем опстаје



Аутоматизација

- Неки сегменти посла у вези са пречишћавањем схеме могу да се аутоматизују
 - установљавање ФЗ на основу узорка података
 - израчунавање затворења скупа ФЗ
 - на основу почетног скупа “претпостављених” ФЗ
 - израчунавање скупа зависних атрибута
 - проналажење кључева кандидата
 - рачуна се скуп зависних атрибута за сваки подскуп атрибута релације
 - X је наткључ ако је $X^+ = Attr(R)$



Затворење скупа атрибута

- Алгоритам “Затворење” из књиге ГПЛ
 - омогућава да се затворење скупа атрибута израчуна итеративно
 - Поступак се понавља све док последња итерација не повећа скуп
- У основи може и да се мало оптимизује
 - Након што се нека ФЗ употреби, може да се искључи из разматраног скупа

Алгоритам ZATVORENJE

Улаз: скуп атрибута X и скуп FZ F ;

Иzlaz: затворење од X над F ;

ZATVORENJE(X, F)

BEGIN

$stari := \emptyset; novi := X;$

 WHILE $novi \neq stari$ DO

 BEGIN

$stari := novi;$

 FOR svaka FZ $W \rightarrow Z$ u F DO

 IF $novi \supseteq W$ THEN $novi := novi \cup Z$

 END;

 RETURN($novi$)

END.

Алгоритам ZATVORENJE

Улаз: скуп атрибута X и скуп FZ F ;

Иzlaz: затворење од X над F ;

ZATVORENJE(X, F)

BEGIN

$stari := \emptyset; novi := X;$

 WHILE $novi \neq stari$ DO

 BEGIN

$stari := novi;$

 FOR svaka FZ $W \rightarrow Z$ u F DO

 IF $novi \supseteq W$ THEN $novi := novi \cup Z$

 END;

 RETURN($novi$)

END.

- Нека су дати:
 - почетни скуп атрибута $X = \{A, E\}$
 - скуп зависности $F = \{A \rightarrow D, AB \rightarrow E, BI \rightarrow E, CD \rightarrow I, E \rightarrow C\}$
- На почетку:
 - $novi = \{A, E\}$
- Након више итерација скуп постаје:
 - $novi = \{A, E, D, C\}$
 - $novi = \{A, E, D, C, I\}$
 - $novi = \{A, E, D, C, I\}$

Алгоритам ZATVORENJE

Улаз: скуп атрибута X и скуп FZ F ;

Иzlaz: затворење од X над F ;

ZATVORENJE(X, F)

BEGIN

$stari := \emptyset; novi := X;$

 WHILE $novi \neq stari$ DO

 BEGIN

$stari := novi;$

 FOR svaka FZ $W \rightarrow Z$ u F DO

 IF $novi \supseteq W$ THEN $novi := novi \cup Z$

 END;

 RETURN($novi$)

END.

- Нека су дати:
 - почетни скуп атрибута $X = \{A, B\}$
 - скуп зависности $F = \{A \rightarrow D, AB \rightarrow DE, CE \rightarrow G, E \rightarrow H\}$
- На почетку:
 - $novi = \{A, B\}$
- Већ након прве итерација скуп постаје:
 - $novi = \{A, B, D, E, H\}$

Затворење скупа ФЗ



- Своди се на једноставније алгоритме:
 - Проверавање да ли је изабрана ФЗ $X \rightarrow Y$ изводива из датог скупа ФЗ F :
 - $\text{CLAN}(X \rightarrow Y, F) = Y \subseteq \text{ZATVORENJE}(X, F)$
 - Назив алгоритма је преузет из књиге ГПЛ. Можда је бољи назив *ИЗВОДИВА*.
 - Проверавање да ли је изабран скуп ФЗ G изводив из датог скупа ФЗ F :
 - $\text{IZVODIV}(F, G) = (\forall f \in G) \text{CLAN}(f, F)$
 - Проверавање да ли су два скупа ФЗ међусобно еквивалентна, у смислу да имају исто затворење:
 - $\text{EKVIV}(F, G) = \text{IZVODIV}(F, G) \wedge \text{IZVODIV}(G, F)$



Минималност скупа ФЗ

- Скуп ФЗ је минималан (нередундантан) ако нема прави подскуп који му је еквивалентан

Алгоритам NEREDUND

Улаз: скуп ФЗ G ;

Иzlaz: neredundantno pokrивање F скупа G ;

NEREDUND(G)

BEGIN

$F := G$;

FOR svaka ФЗ $X \rightarrow Y$ из G DO

IF ČLAN($F \setminus \{X \rightarrow Y\}, X \rightarrow Y$) THEN $F := F \setminus \{X \rightarrow Y\}$;

RETURN (F)

END.



Каноничко покривање скупа ФЗ

- Скуп ФЗ је канонички ако свака ФЗ има облик $X \rightarrow A$, где је X скуп атрибута, а A један атрибут
- Каноничко покривање скупа ФЗ је нередундантно покривање скупа ФЗ у коме је
 - свака ФЗ каноничка
 - на левој страни ФЗ не постоје небитни атрибути (они чије уклањање не мења затворење)
- У општем случају, каноничко покривање скупа ФЗ се рачуна у три корака:
 - свака ФЗ облика $X \rightarrow Y$ се замени скупом ФЗ $\{X \rightarrow A \mid A \in Y\}$
 - за сваку ФЗ $X \rightarrow A$ из добијеног скупа ФЗ, из леве стране се уклоне сви небитни атрибути
 - израчуна се нередундантно покривање добијеног скупа ФЗ



Резиме

- У оквиру пречишћавања схеме обрадили смо:
 - зависности
 - функционалне зависности
 - вишезначне зависности
 - зависности спајања
 - нормалне форме
 - основне: 1НФ, 2НФ, 3НФ, НФЕК, БКНФ
 - више: 4НФ, НФЕТ, НФБР, НФНК, 5НФ, НФДК, 6НФ
 - поступак нормализације
 - неке алгоритме

Литература за тему

- Teorey, Lightstone, Nadeau, Jagadish, **Database Modeling and Design**, 5.ed, Elsevier, 2011.
- Watt, Eng, **Database Design**, 2.ed, Open Edition, 2014.
- Гордана Павловић-Лажетић, **Увод у релационе базе података**, 2.изд. Математички факултет, 1999.
 - доступно онлајн: <http://poincare.matf.bg.ac.rs/~gordana/urbp-2016.htm>
- Ramakrishnan, Gehrke, **Database Management Systems**, 2.ed, 2000.
- Darween, Date, Fagin, **A Normal Form for Preventing Redundant Tuples in Relational Databases**, ICDT '12 Proceedings, 2012.