

Нормализација

Ненад Митић

Математички факултет
nenad.mitic@matf.bg.ac.rs

Пројектовање база података

- Логичко пројектовање/физичко пројектовање базе
 - Логичко пројектовање базе
 - Нормализација - коришћење идеја о нормализацији ради разбијања "великих" у "мале" релације
 - Семантичко моделирање - употреба модела ентитета и односа ради формирања "великих" релација

Нормализација - увод

- Нормализација је процес замене релација скупом релација које су у **погоднијем облику**
 - Сврха нормализације је **избегавање редундантности и поједињих аномалија ажурирања**

Нормализација - увод

У процесу нормализације оператор пројекције се више пута примењује на дату релацију на такав начин да спајањем пројекција може да се дође до почетне релације. На тај начин, процес нормализације је реверзибилан и чува информације, тј. увек је могуће да се узме излаз из процеса и преслика унатраг до улаза

Нормалне форме

Първа нормална форма (1NF)

Друга нормална форма (2NF)

Трећа нормална форма (3NF)

Бојс-Кодова нормална форма (BCNF)

Четврта нормална форма (4NF)

Пета нормална форма (5NF)

Аномалије

Пример: релација `PredmetIspit` која је добијена 'мешањем' релација `Predmet` и `Ispit`. Релација садржи атрибуте `{Id_predmeta, Sifra, Naziv, Bodovi, Indeks, Ocena}`

Примарни кључ је {Id_predmeta, Indeks}, а важи и додатна
ФЗ *Naziv*—>*Bodovi*.

ID_PREDMETA	SIFRA	NAZIV	BODOVI	Indeks	Ocena
1001	M111	Analiza 1	6	20100021	7
1002	M112	Analiza 2	6	20100021	8
1001	M111	Analiza 1	6	20100022	7
1021	M131	Geometrija	6	20100021	8
1101	M105	Diskrete strukture 1	5	20100021	6
1101	M105	Diskrete strukture 1	5	20100023	6
2002	P102	Programiranje 2	8	20100024	9
2002	P102	Programiranje 2	8	20100025	9
2003	P103	Objektno orijentisano programiranje	8	20100021	7
2004	P104	Algoritmi i strukture podataka	8	20100021	6
4001	R101	Uvod u organizaciju racunara	7	20100021	10
4002	R102	Uvod u Web i Internet tehnologije	7	20100021	10

Аномалије

Постоји редундантност: сваки `Id_predmeta` 1001 као назив показује *Analiza 1*. Такође, свака *Analiza 1* за број бодова има вредност 5, итд.

ID_PREDMETA	SIFRA	NAZIV	BODOVI	Indeks	Ocena
1001	M111	Analiza 1	6	20100021	7
1001	M111	Analiza 1	6	20100022	7

Аномалије

Аномалије ажурирања:

- Уношење: не може једноставно да се унесе податак да неки Id одговара поједином називу предмета док неко није полагао тај предмет (јер не постоји одговарајући број индекса који је део п.к)
- Ако се избриши сви подаци за Id_predmeta, такође се бриши и подаци о називу тог предмета, његовој шифри, броју бодова,
...
- Пошто се исти назив јавља на више места, то може да доведе до проблема при промени назива

Због тога се врши декомпозиција (преко пројекција)
оваквих релација

Прва нормална форма

Релвар је у 1NF ако и само ако у свакој важећој вредности тог релвар-а свака торка садржи тачно једну вредност за сваки атрибут

Друга нормална форма

Релвар је у 2NF ако и само ако је сваки некључни атрибут нередуцибилно зависан од примарног кључа

Трећа нормална форма

Релвар је у 3НФ ако и само ако је у 2НФ и сваки некључни атрибут је нетранзитивно зависан од примарног кључа

Претходне две дефиниције подразумевају постојање само једног кандидата за кључ који је истовремено и примарни кључ. Последица: некључни атрибути су узајамно независни

Друга нормална форма - дефиниција преко ФЗ

Релвар R је у 2NF ако и само ако за сваку нетривијалну $\Phi_3 X \rightarrow Y$ из R важи:

- X је суперкључ, или
- Y је део кључа, или
- X није прави подскуп кључа

Трећа нормална форма - дефиниција преко ФЗ

Релвар R је у 3NF ако и само ако за сваку нетривијалну ФЗ: $A_1A_2...A_n \rightarrow B_1B_2...B_m$ из R важи

- $\{A_1A_2...A_n\}$ је суперкључ, или
- сваки атрибут $B_1B_2...B_m \notin \{A_1A_2...A_n\}$ је елемент неког (не обавезно истог) кандидата за кључ

(или је лева страна суперкључ или су атрибути на десној страни ФЗ део кључа)

Недостаци треће нормалне форме

Кодова оригинална дефиниција ЗНФ није узимала у обзир случајеве када

- 1 релација има више од једног кандидата за кључ
 - 2 је кандидат за кључ је композитан
 - 3 се композитни кандидати за кључеве се преклапају

Ови случајеви су обухваћени Бојс-Кодовом нормалном формом

Недостаци треће нормалне форме (наставак)

Пример: релација SPN
(Student-Predmet-Nastavnik)

S	P	N
Lazić	matematika	Petrović
Lazić	računarstvo	Marković
Perić	matematika	Petrović
Perić	računarstvo	Marković

Значење торке: студент S слуша предмет P код наставника N.

Недостаци треће нормалне форме (наставак)

Претпоставке:

- ① за сваки предмет сваки студент слуша наставу само код једног наставника, тј.
 $\{S,P\} \rightarrow N$
- ② Кандидати за кључ су $\{S,P\}$ и $\{S,N\}$
- ③ ФЗ $\{S,P\} \rightarrow N$
- ④ Сваки наставник предаје само један предмет,
тј. $N \rightarrow P$

Бојс-Кодова нормална форма

Релвар је у BCNF ако и само ако свака нетривијална лево-нередуцибилна ФЗ има кандидат за кључ као своју леву страну

Алтернативно: Релвар је у BCNF ако и само ако свака нетривијална ФЗ има суперкључ као своју леву страну

Пример редукције у BCNF

Нека су називи предмета јединствени. Тада у релацији $PredmetDosiye\{Id_predmeta, Naziv, Indeks, Ocena\}$ важе следеће ФЗ

На основу кандидата за кључ

- $\{Id_predmeta, Indeks\} \rightarrow Ocena$
- $\{Naziv, Indeks\} \rightarrow Ocena$

На основу јединствености назива предмета

- $Id_predmeta \rightarrow Naziv$
- $Naziv \rightarrow Id_predmeta$

Пример редукције у BCNF

Id_predmeta и Naziv су на левој страни $\Phi_3 FZ$, али нису кандидати за кључ

⇒ PredmetDosije није у BCNF

Решење: разбијање PredmetDosije на две релације:

- ① $IN\{Id_predmeta, Naziv\}$ и
 $PD\{Id_predmeta, Indeks, Ocena\}$, или
- ② $IN\{Id_predmeta, Naziv\}$ и
 $PD\{Naziv, Indeks, Ocena\}$

Могуће редукције

У релацији

Dosije {Indeks, Jmbg, Datum_rođenja, ...}

кандидати за кључ су Indeks и Jmbg. Нека важе следеће
ФЗ:

- Indeks \rightarrow Jmbg
- Jmbg \rightarrow Datum_rođenja
- ... \rightarrow ...

Релација није у 3NF јер постоји транзитивна зависност
 $Indeks \rightarrow Datum_rođenja$.

Могуће редукције (наставак)

Разбијање релације Dosije

Dosije_1a {Indeks, Jmbg, ...}

Dosije_1b {Jmbg, Datum_rođenja, ...}

Dosije_2a {Indeks, Jmbg, ...}

Dosije_2b {Indeks, Datum_rođenja, ...}

Dosije_3a {Jmbg, Datum_rođenja, ...}

Dosije_3b {Indeks, Datum_rođenja, ...}

Могуће редукције (наставак)

- Сва три паре релација јесу у 3NF и у BCNF
 - Декомпозиције [1] и [2] не доводе до губитка информација
 - При декомпозицији [2] јавња се аномалија при уносу

\implies коректно разбијање је као у случају [1].

Могуће редукције (наставак)

Пожељне особине при редукцији:

- 1 Елиминација аномалија
 - 2 Могућност реконструкције информација - да ли почетни облик релације може да се реконструише из нових релација које су добијене декомпозицијом?
 - 3 Очување ФЗ. Ако се ФЗ пројектују на декомпоноване релације, да ли ће реконструисана оригинална релација (добијена спајањем релација које су резултат декомпозиције) задовољавати оригиналне ФЗ?

Могуће редукције (наставак)

Хитова (енг. *Heath*) теорема:

Нека је A скуп атрибута релације R и нека су $X \subseteq A, Y \subseteq A, Z \subseteq A$ тако да важи $X \cup Y \cup Z = A$. Нека $XY = X \cup Y$, и $XZ = X \cup Z$.

Тада, ако у R важи $\Phi 3 X \rightarrow Y$, тада је R једнако спајању пројекција на XY и XZ

Делимичан одговор - ако се декомпозиција врши према ФЗ тада се информације чувају, али ако се врши на други начин не даје се одговор да ли су информације очуване или не



Могуће редукције (наставак)

Правила:

- ① Све ФЗ полазног скупа морају да буду очуване (директно или могућим извођењем из скупа релација добијених декомпозицијом)
- ② Ако у новодобијеним пројекцијама насталим разбијањем основне релације постоји заједнички атрибут, он мора да буде кључ у бар једној од новодобијених релација

Могуће редукције (наставак)

Алгоритам за декомпозицију релације R са скупом $\Phi_3 S$ у BCNF

- 1 Проверити да ли је релација R у BCNF. Ако јесте, процес је завршен и R је решење
- 2 Ако $\Phi_3 X \rightarrow Y$ нарушава BCNF тада за нове релације узети $R_1 = \{X\}^+$ и $R_2 = \{X\} \cup (R - \{X\}^+)$
- 3 Одредити скупове S_1 и S_2 Φ_3 за релације R_1 и R_2 користећи алгоритам за одређивање скупа Φ_3 у пројекцији релације
- 4 Применити рекурзивно алгоритам на релације R_1 и R_2
- 5 Као решење узети унију свих декомпозиција

Вишезначне зависности

Релација PNU (Predmet-Nastavnik-Udžbenik):

P	N	U
matematika	{Petrović, Marković}	{Analiza, Linearna algebra}
računarstvo	Petrović	{Linearna algebra, Uvod u programiranje, Strukture podataka}

Значење торке: предмет P може да предаје било који наставник N и да користи било који уџбеник U

Вишезначне зависности (наставак)

Предпоставке:

- 1 за дати курс постоји произвољан број наставника и уџбеника
 - 2 наставници и текстови су независни
 - 3 наставник или текст може да се придружи било ком курсу

У овој релацији не постоје ФЗ

Вишезначне зависности (наставак)

- Примарни кључ: {Predmet, Nastavnik, Udžbenik}
- Релавар је у 1NF, 2NF, 3NF и BCNF
- Релвар РНУ поседује редундантост
 - ако обе торке (p, n_1, u_1) и (p, n_2, u_2) постоје тада морају да постоје и торке (p, n_1, u_2) и (p, n_2, u_1)
- Последица је аномалију при ажурирању: да би се унео податак да нови наставник предаје математику морају да се унесу две торке, по једна за сваки уџбеник

Вишезначне залежності (наставак)

Интуитивно може да се изврши декомпозиција

- {predmet, nastavnik} и
 - {predmet, udžbenik}

Декомпозиција не следи из ФЗ (којих и нема) већ из постојања вишезначних зависности

Дефиниција вишезначних зависности

Дефиниција:

Нека је R релвар и нека су A , B и C подскупови атрибута од R . Каже се да је **вишезначно зависно** (**VZ**) од A , у означи $A \rightarrow\!\!\!-\!\!\!-\! B$, ако и само ако у свакој могућој важећој вредности од R , скуп вредности B који се упарује са паром (вредност A , вредности C) зависи једино од вредности A и независан је од вредости C .

Четврта нормална форма

Релвар R је у 4NF ако и само ако је у BCNF и сваки пут када постоје подскупови A и B атрибута од R такви да је задовољена нетривијална вишезначна зависност $A \rightarrow\rightarrow B$, тада су сви атрибути од R такође функционално зависни од A

Примедба: $VZ A \rightarrow\rightarrow B$ је тривијална ако је или A надскуп од B или је $A \cap B$ садржи све атрибуте од R

Зависност спајања

Нека је R релвар и нека су A, B, \dots, Z подскупови атрибута од R . Тада R задовољава **зависност спајања (ZS)** * $\{A, B, \dots, Z\}$ ако и само ако је R у $4NF$ и свака могућа важећа вредност у R је једнака спајању њених пројекција на A, B, \dots, Z

Пета нормална форма

Релвар R је у 5NF (пројекција-спајање-NF) ако и само ако R је у 4NF и свака нетривијална зависност спајања која важи у R је последица кандидата за кључ у R , где

- Зависност спајања $*\{A, B, \dots, Z\}$ у R је тривијална ако је најмање један од A, B, \dots, Z скуп сви атрибута R
- Зависност спајања $*\{A, B, \dots, Z\}$ у R је последица кандидата за кључ релвара R ако је сваки од A, B, \dots, Z надкључ за R

Теорема: Ако је релација у BCNF и нема композитне кључеве, тада је она и у 5NF

Редефиниције НФ

Релациона променљива R је

- 1) нормализована ако је у 1NF (у релационом моделу сви релвари су у 1NF)
- 2) у 2NF ако за сваки кључ K у R и сваки некључни атрибут A из $R \text{ FZ } K \rightarrow \{Y\}$ која важи у R је нередуцибилна (тј. K је минимални кључ)
- 3) у 3NF ако свака нетривијална $\Phi3 X \rightarrow Y$ која важи у R или је X суперкључ или је Y подскуп кључа
- 4) у BCNF ако у свакој нетривијалној $\Phi3 X \rightarrow Y$ која важи у R , X је суперкључ

Процес нормализације

- 1) Узети пројекције оригиналног релвара у 1NF ради елиминисања ФЗ које нису нередуцибилне. Добијени скуп релвара је у 2NF.
- 2) Узети пројекције релвара у 2NF ради елиминисања транзитивних зависности. Добијени скуп релвара је у 3NF.
- 3) Узети пројекције релвара у 3NF ради елиминисања преосталих ФЗ у којима на левој страни није кандидат за кључ. Добијени скуп релвара је у BCNF

Процес нормализације

- 4) Узети пројекције релвара у BCNF ради елиминисања VZ које нису и ФЗ. Добијени скуп релвара је у 4NF.
- 5) Узети пројекције релвара који су у 4NF и елиминисати ZS које не следе из кандидата за кључ(еве). Добијени скуп релвара је у 5NF.

Денормализација

У пракси се често не спроводи пунा нормализација због добрих перформанси

- пуну нормализација доводи до великог броја логички раздвојених релвар-а
- велики број раздвојених релвар-а значи велики број раздвојених датотека у којима се чувају
- велики број датотека значи велики број У/И операција

У пракси се нормализација најчешће спроводи до BCNF

Пример 1

Нека је дат релвар $R = \{A, B, C, D\}$ и скуп F ФЗ:

- 1) $AB \rightarrow C$
- 2) $C \rightarrow D$
- 3) $D \rightarrow A$

- 1 Навести неке нетривијалне ФЗ које могу да буду изведене из F
- 2 Одредити кандидате за кључ релвара R и испитати да ли је релација у 3NF
- 3 Навести све ФЗ које спречавају да релавар R буде у BCNF
- 4 Декомпоновати R тако да добијене релације буду у BCNF
- 5 Одредити ФЗ које важе за новодобијене релације

Пример 1 - наставак

Неке нетривијалне ФЗ су:

- | | |
|--------------------------|----------------------------|
| 1) $C \rightarrow ACD$ | 6) $BD \rightarrow ABCD$ |
| 2) $D \rightarrow AD$ | 7) $CD \rightarrow ACD$ |
| 3) $AB \rightarrow ABCD$ | 8) $ABC \rightarrow ABCD$ |
| 4) $AC \rightarrow ACD$ | 9) $ABD \rightarrow ABCD$ |
| 5) $BC \rightarrow ABCD$ | 10) $BCD \rightarrow ABCD$ |

Пример 1 - наставак

Из затворења скупа атрибута

- | | |
|--------------------------|---------------------------|
| 1) $\{A\}^+ = \{A\}$ | 7) $\{AD\}^+ = \{AD\}$ |
| 2) $\{B\}^+ = \{B\}$ | 8) $\{BC\}^+ = \{ABCD\}$ |
| 3) $\{C\}^+ = \{ACD\}$ | 9) $\{BD\}^+ = \{ABCD\}$ |
| 4) $\{D\}^+ = \{AD\}$ | 10) $\{CD\}^+ = \{ACD\}$ |
| 5) $\{AB\}^+ = \{ABCD\}$ | 11) $\{ACD\}^+ = \{ACD\}$ |
| 6) $\{AC\}^+ = \{ACD\}$ | |

добија се да су кандидати за кључ AB, BC и BD (остале комбинације дужине 3 и 4 су суперкључеви)

Пример 1 - наставак

Упутство: испитивање да ли је релација у 3НФ.
Користити дефиницију припадности 3НФ у
 зависности од функционалних зависности

Пример 1 - наставак

Ф3 које онемогућавају да релација R буде у БЦНФ су

- 1) $C \rightarrow D$ (FZ 2) из почетног скупа
 - 2) $D \rightarrow A$ (FZ 3) из почетног скупа

јер атрибути који се налазе на њиховим левим странама нису кандидати за кључ

Пример 1 - наставак

Неформалан приступ:

- 1) Ако се релација $R = \{A, B, C, D\}$ разбије на основу ФЗ 2) добија се $R_1 = \{C, D\}$ и $R_x = \{A, B, C\}$ (искључује се атрибут који је на десној страни ФЗ) која се затим разбија на основу ФЗ $C \rightarrow A$ (транзитивност!), тако да је крајњи резултат $R_1 = \{C, D\}$ и $R_2 = \{C, A\}$ и $R_3 = \{B, C\}$
- 2) Ако се релација $R = \{A, B, C, D\}$ разбије на основу ФЗ 3) добија се $R_1 = \{D, A\}$ и $R_x = \{B, C, D\}$ (искључује се атрибут који је на десној страни ФЗ) која се затим разбија на основу ФЗ 2), тако да је крајњи резултат $R_1 = \{D, A\}$ и $R_2 = \{C, D\}$ и $R_3 = \{B, C\}$

Домаћи задатак:

- одредити према алгоритму (формално) разбијање релација у БЦНФ
- одредити ФЗ које важе за новодобијене релације

Пример 2

Нека је дат релвар $R = \{A, B, C, D, E, F\}$ и скуп ФЗ:

- 1) $AB \rightarrow D$
- 2) $B \rightarrow C$
- 3) $AE \rightarrow B$
- 4) $A \rightarrow D$
- 5) $D \rightarrow EF$

Трансформисати релацију R тако да новодобијена релација буде у BCNF.

Упутство: одредити кандидате за кључ за нередуциблни скуп ФЗ.

Пример 2 - решење

Нередуцибилни скуп Φ_3 је

- 1) $B \rightarrow C$
- 2) $A \rightarrow B$
- 3) $A \rightarrow D$
- 4) $D \rightarrow EF$

Важи $\{A\}^+ = \{ABCDEF\}$ па преостале 2 Φ_3 нарушавају БЦНФ. Разбијањем

- 1) Iz $B \rightarrow C$, како је $\{B\}^+ = \{B, C\}$ добија се $R1(B, C)$,
 $R1a\{A, B, D, E, F\}$
- 2) Iz $D \rightarrow EF$, како је $\{D\}^+ = \{D, E, F\}$ добија се $R2(D, E, F)$,
 $R3\{A, B, D\}$

BCNF чине релације R_1 , R_2 , R_3