Увод у релационе базе података

11а

Саша Малков
Универзитет у Београду
Математички факултет
2023/2024

[PM13]
Увод у РБП
Саша Малков

Тема 7.5

Логичко пројектовање - Вишезначне зависности -

Ограничења функционалних зависности

- Као што им само име каже,
  ФЗ описују ***функционалне*** зависности
  - 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 сасвим 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 школског 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 општијих 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 $\boldsymbol{X \rightarrow\rightarrow Y}$, акко 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->>Y***
- Неформално речено, за сваку изабрану вредност атрибута из X, вредности из Y су независне од вредности из Z, тј. зависе само од X.
- Т.ј. ако су X, Y и Z дисјунктни скупови атрибута који покривају $Attr(R)$, онда важи ВЗ $X \rightarrow\rightarrow Y$ акко за сваку фиксирану вредност Xc рестрикција по услову $X=Xc$ && $Z=Zc$ има исту вредност, без обзира на изабрано Zc.

Вишезначне зависности (пример)

- На претходном примеру можемо да уочимо вишезначне зависности:
  - {Наслов} →→ {Глумац}
  - {Наслов} →→ {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 $W = Attr(R) \setminus XY$...
- ***Комплементарност***
  - Ако важи $X \to\to Y$ онда важи и $X \to\to W$
- ***Рефлексивност***
  - Ако важи $Y \subseteq X$, онда важи и $X \to\to Y$
- ***Проширивост***
  - Ако важи $X \to\to Y$ онда важи и $XZ \to\to YZ$
- ***Транзитивност***
  - Ако на R важи $X \to\to Y$ и $Y \to\to Z$ онда важи и $X \to\to Z \setminus Y$

Веза вишезначних зависности и ФЗ

- ***Конвертибилност***
  - Ако на R важи $X \to Y$ онда важи и $X \to\to Y$
  - Т.ј. ФЗ су спец.случај ВЗ
- ***Интерактивност***
  - Ако на R важи $X \to\to Y$ и $XY \to Z$, онда важи и $X \to Z \setminus Y$
    - 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=Attr(R)\XY.
    Онда на релацији R важи једнакост $R = R[XY] * R[XZ]$ акко важи ВЗ $X \to\to Y$
- Еквивалентно тврђење:
  - Релација R се може потпуно декомпоновати на $R[W]$ и $R[Q]$ акко важи ВЗ: $W \cap Q \to\to W \setminus Q$
    (акко важи ВЗ: $W \cap Q \to\to Q \setminus W$)

ВЗ и декомпозиција (2)

- Веома је важно да постоји еквиваленција (акко) између ВЗ и потпуне декомпозиције једне релације на две релације
- Одатле непосредно следи да је установљавање ВЗ довољан и неопходан услов за потпуну декомпозицију једне релације на две мање релације

Тривијалне ВЗ

- Ако је X непразан скуп атрибута, онда кажемо да су ВЗ $X \rightarrow\rightarrow \emptyset$ и $X \rightarrow\rightarrow$ Attr(R)\X ***тривијалне***
  - наравно, ако важе

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,... X_n скупови атрибута релације R, чија је унија управо $Attr(R)$. У релацији R важи ***зависност спајања***, у ознаци *{X_1, X_2,... X_n} акко је:
  R = R[X1] * R[X2] * ... * R[Xn]
- 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 $*\{X_1, X_2, \ldots X_n\}$ је тривијална акко за неки од скупова атрибута X_i важи $X_i = Attr(\mathrm{R})$.

[PM13]

Увод у РБП

Саша Малков

Тема 7.6

Логичко пројектовање - 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 на основу сложенијих облика зависности
- Већина је уређена по сложености (виша имплицира нижу):
  - 4.НФ
  - НФЕТ – нормална форма есенцијалних торки
  - НФБР – нормална форма без редунданси
  - НФНК – нормална форма наткључева
  - 5.НФ
  - НФДК – нормална форма домена и кључа
- 6.НФ не може да се “пореЂа”

4. нормална форма

- Релација R је у 4.НФ акко за сваку вишезначну зависност $X \to\to Y$ 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)

- Нека имамо релацију
  - *Понуда(__продавница__, __производ__, __град__)*
- и нека важи правило
  - { *продавница* } →→ { *производ* }
    - тј. исти скуп производа се нуди независно од града
    - тј. у сваком граду се нуде сви производи које продавница нуди
  - наравно, онда мора да важи и { *продавница* } →→ { *град* }
- 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 при пажљивом посматрању
  - зато се често заобилази у настави база података
    - као и вишезначне зависности
  - очекује се да ће проблем бити уочен и решен интуитивно
  - али ипак се повремено појављује у продукционим базама података
- Незадовољавање 4.НФ може да доведе до драстичног нарушавања перформанси, због значајног повећавања броја редова

Нормална форма есенцијалних торки

- Релација R задовољава НФЕТ акко свака торка ове релације мора да буде *есенцијална*.
  - Торка је *есенцијална* акко није ни потпуно ни делимично редундантна.
  - Торка t је потпуно редундантна у релацији R акко 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 акко у истој релацији постоји ФЗ X->A и торка q ($t \neq q$) таква да су им пројекције на X једнаке: $t[X] = q[X]$.
- Важи (теорема):
  - Нека је релација R одређена само функционалним зависностима и зависностима спајања. Релација R је у НФЕТ акко је у БКНФ и свака зависност спајања има неку компоненту која је наткључ.
- (Darween et al., 2012)

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→ J == "снабдевач испоручује неки део за највише један пројекат"
  - она је у 4.НФ
    - нема других ФЗ
    - има само тривијалне ВЗ
    - нетривијална ЗС не може да се изведе из ФЗ
  - 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 елиминише редундантне торке
    - могућа је редундантност на нивоу атрибута, коју превазилази НФБР
- Представља теоријски резултат којим се показује да за потпуну елиминацију редундантности не мора да се оствари 5.НФ

Нормална форма без редунданси

- Релација *R* задовољава НФБР акко 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 не може да се у релацији замени само та једна вредност а да релација и даље остане исправна
- (Darween et al., 2012)

Нормална форма наткључева

- Релација *R* задовољава НФНК акко 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)

5. нормална форма

- Релација *R* је у 5.НФ акко је свака нетривијална зависност спајања имплицирана кључевима кандидатима.
  - Зависност спајања је имплицирана кључевима акко свака компонента зависности спајања представља натккључ релације
  - Позната је и под именом Нормална форма пројективног спајања (PJNF).
- Пример релације у 5.НФ је свака релација у којој су некључни атрибути (сваки за себе) функционално зависни од кључа
  - нпр. *Студент(<u>индекс</u>, име, презиме)*
  - нетривијална ЗС је *{ { индекс, име } , { индекс, презиме } }

5. нормална форма (пример 1)

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

5. нормална форма (пример 2)

- Нека поново имамо пример са продавницама:
  - *Понуда(**<u>продавница</u>**, **<u>производ</u>**, **<u>град</u>**)*
- 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
- Релација није у 5.НФ – постоји декомпозиција (зависност спајања):
  - *({ *продавница, производ* }, { *продавница, град* }, { *производ, град* })
- Проблем се решава наведеном декомпозицијом на три релације
- Не постоји декомпозиција на две релације
  - иначе би постојала ВЗ

Нормална форма домена и кључа

- Релација задовољава ***НФДК*** акко 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 акко не задовољава ниједну зависност спајања.
  - тј. акко не постоји потпуна декомпозиција
  - тј. акко 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 примарни кључ

[PM13]

Увод у РБП

Саша Малков

Тема 7.2

Логичко пројектовање - Пречишћавање схеме -

...наставак...

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)

- Често се као крајњи циљ нормализације поставља БКНФ или 4.НФ
  - али не мора да буде тако
- 4.НФ и друге више НФ углавном могу да се интуитивно задовоље
  - ако се пажљиво анализирају схема и функционалне зависности
  - ако пројектанти имају довољно искуства

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 натկључ
    - 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\A*
  - 2. Ако *R\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:
  - ***ПријаваИспита***(*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 *ПријаваИспита* 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 акко 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

Algoritam ZATVORENJE
Ulaz: skup atributa X i skup FZ F;
Izlaz: zatvorenje od X nad 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.

Algoritam ZATVORENJE
Ulaz: skup atributa X i skup FZ F;
Izlaz: zatvorenje od X nad 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 = $\{A, E, D, C, I\}$

Algoritam ZATVORENJE
Ulaz: skup atributa X i skup FZ F;
Izlaz: zatvorenje od X nad 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:
    - ČLAN($X \rightarrow Y$, F) = $Y \subseteq$ 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:
    - IZVODIV(F, G) = $(\forall f \in G)$ ČLAN(f, F)
  - Проверавање да ли су два скупа ФЗ међусобно еквивалентна, у смислу да имају исто затворење:
    - EKVIV(F, G) = IZVODIV(F, G) $\wedge$ IZVODIV(G, F)

Минималност скупа ФЗ

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

Algoritam NEREDUND
Ulaz: skup FZ G;
Izlaz: neredundantno pokrivanje F skupa G;
NEREDUND(G)
BEGIN
  $F := G$;
  FOR svaka FZ $X \rightarrow Y$ iz 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:
  1. свака ФЗ облика $X \rightarrow Y$ се замени скупом ФЗ $\{ X \rightarrow A \mid A \in Y \}$
  2. за сваку ФЗ $X \rightarrow A$ из добијеног скупа ФЗ,
     из леве стране се уклоне сви небитни атрибути
  3. 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
  - неки алгоритми

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.*