NORMALIZACIJA

2. deo

ZADOVOLJENOST F-NE ZAVISNOSTI

Ako u relaciji R važi funkcionalna zavisnost $X \to Y$, onda se kaže da relacija R *zadovoljava* tu funkcionalnu zavisnost.

Relacija R je dinamičkog sadržaja, što znači da njen sadržaj u raznim momentima može biti različit. *Stanje* r relacije R je njen sadržaj u nekom momentu. Kazaćemo da stanje r relacije R *zadovoljava* FZ $X \to Y$ ako ta funkcionalna zavisnost važi u relaciji R u momemtu koji definiše njeno stanje r. Sada se za relaciju R može reći da zadovoljava FZ $X \to Y$ ako svako njeno stanje r zadovoljava tu FZ.

FUNKCIONALNE (FUNKCIJSKE) ZAVISNOSTI

- $X \rightarrow Y$ je trivijalna akko za važi da je $Y \subseteq X$
- $X \rightarrow Y$ je potpuna akko ne postoji ni jedan pravi podskup Z od X, za koji važi $Z \rightarrow Y$
- $X \rightarrow Y$ je parcijalna akko za $X \rightarrow Y$ postoji pravi podskup Z od X, za koji važi $Z \rightarrow Y$
- $X \rightarrow Y$ je tranzitivna akko za $X \rightarrow Y$ postoji Z različito od X i Y za koje važi $X \rightarrow Z$ i $Z \rightarrow Y$

X, Y, Z - podskupovi skupa atributa relacije R

IZVOĐENJE F-NIH ZAVISOSTI

- Neka je zadat skup *F* funkcionalnih zavisnosti jedne relacije:
  - Kako konstruisati skup svih funkcionalnih zavisnosti koje se mogu izvesti iz *F* po Armstrongovim aksiomama?
  - Da li je zadata FZ logička posledica skupa *F*? (Da li je izvodiva iz F)

ZATVORENJE SKUPA F-IH ZAVISNOSTI

- Neka je F neki skup FZ relacije R. **Zatvorenje od F**, u oznaci F+; jeste najmanji skup FZ koji sadrži F i takav da se primenom Armstrongovih aksioma na F+ ne dobija nijedna FZ koja već nije u F+.

- Primer

  R(A,B,C)

  $F = \{A \rightarrow B, A \rightarrow C\}$

  $F^+ = \{ A \rightarrow A, \quad A \rightarrow B, \quad A \rightarrow C, A,B \rightarrow A, \quad A,B \rightarrow B,$
  $A,B \rightarrow C \;\; A,B \rightarrow A, B \;\; A,B \rightarrow B,C \;\; A,B \rightarrow A,C,$
  $A,B \rightarrow A,B,C, A, C \rightarrow \ldots, A,B,C \rightarrow A,B,C\}$

Da bismo odredili KK…
Najčeće nepotrebno.

ZATVORENJE X NAD F

- Neka je R skup atributa, X neki njegov podskup, a F skup f-ih zavisnosti nad R. **Zatvorenje skupa atributa X nad skupom f-ih zavisnosti F,** u oznaci X^+ ili $(X,F)^+$, čini skup atributa Y koji odgovara desnoj strani zavisnosti $X \rightarrow Y$ u zatvaraču skupa funkcionalnih zavisnosti sa maksimalnim Y.

  $(X, F)^+ := \{A_i \mid X \rightarrow A_i$ *se može izvesti iz* $F\}$

  Neformalno: Maksimalan skup atributa f-no zavisnih od X.

- Kako doći do zatvorenja skupa atributa X nad skupom f-nih zavisnosti F?

```
unused := F
closure := X
do {
  for(Y → Z ∈ unused) {
    if(Y ⊆ closure) {
      unused := unused \ {Y → Z}
      closure := closure ∪ Z
    }
  }
} while(unused and closure did not change)
return closure
```

ZATVORENJE X NAD F

$F = \{ \quad \{A\} \rightarrow \{B, C\}, \quad \{E\} \rightarrow \{C, G\},$
$\{B\} \rightarrow \{E\}, \quad \{C, D\} \rightarrow \{E, G\} \}$

Odrediti $Y = (\{A,B\},F)^+$ / $(A,B)^+$

Neka je $X = \{A,B\}$, $Y = (\{A,B\},F)^+$

Algoritam:

$Y = \{A,B\}$

dodajemo C, zbog $\{A\} \rightarrow \{B, C\}$ $\quad Y = \{A,B,C\}$

dodajemo E, zbog $\{B\} \rightarrow \{E\}$ $\quad Y = \{A,B,C,E\}$

dodajemo G, zbog $\{E\} \rightarrow \{C,G\}$ $\quad Y = \{A,B,C,E,G\}$

$(\{A,B\},F)^+ = \{A,B,C,E,G\}$

ČEMU SLUŽI $(X,F)^+$

- Za dat skup f-ih zavisnosti F možemo odrediti da li $X \to Y$ pripada F+

  Treba proveriti da li $Y \subseteq (X, F)^+$

- Da li je dati skup atributa X nadključ relacije R

  Treba proveriti da li $(X, F)^+$ = skup svih atributa relacije R

POKRIVANJE SKUPA FZ

Definicija 8.3 Dva skupa FZ F i F' relacije R su *ekvivalentna* (u oznaci $F \equiv F'$) ako je $F^+ = F'^+$. Ako je $F \equiv F'$, onda je F *pokrivanje za* F' (F' je *pokrivanje za* F).

Definicija 8.4 Skup FZ F je *neredundantan* ako ne sadrži pravi podskup $F' \subset F$ takav da je $F' \equiv F$; u suprotnom je skup F *redundantan*. Skup F je *neredundantno pokrivanje* skupa FZ G ako je F pokrivanje skupa G i F je neredundantan skup FZ.

Definicija 8.6 *Kanoničko pokrivanje* skupa FZ G je neredundantno pokrivanje F koje zadovoljava sledeća dva uslova:

a) svaka FZ u skupu F je oblika $X \rightarrow A$, gde je A jedan atribut;

b) leva strana svake FZ iz F oslobođena je *nebitnih* atributa, tj. atributa čije udaljenje iz leve strane te FZ ne menja zatvorenje skupa FZ G.

KANONIČKO POKRIVANJE SKUPA FZ

Definicija 8.6 *Kanoničko pokrivanje* skupa FZ G je neredundantno pokrivanje F koje zadovoljava sledeća dva uslova:

a) svaka FZ u skupu F je oblika $X \to A$, gde je A jedan atribut;

b) leva strana svake FZ iz F oslobođena je *nebitnih* atributa, tj. atributa čije udaljenje iz leve strane te FZ ne menja zatvorenje skupa FZ G.

KANONIČKO POKRIVANJE SKUPA FZ

$$
\begin{array}{lcl|c|lcl|c|lcl|c|lcl|c|lcl}
A \to BC & & & & A \to B & & & & A \to B & & & & A \to B & & & & A \to B \\
B \to C & & & & A \to C & & & & A \to C & & & & \cancel{A \to C} & & & & \cancel{A \to C} \\
A \to B & \longrightarrow & & & B \to C & \longrightarrow & & & B \to C & \longrightarrow & & & B \to C & \longrightarrow & & & B \to C \\
AB \to C & & & & \cancel{A \to B} & & & & \cancel{A \to B} & & & & \cancel{A \to B} & & & & \cancel{A \to B} \\
AC \to D & & & & AB \to C & & & & AB \to C & & & & \cancel{AB \to C} & & & & \cancel{AB \to C} \\
 & & & & AC \to D & & & & \cancel{AC \to D} & & & & \cancel{AC \to D} & & & & \cancel{AC \to D} \\
 & & & & & & & & A \to D & & & & A \to D & & & & A \to D
\end{array}
$$

NORMALIZACIJA - NASTAVAK

- Postoje sledeće dve varijante vertikalne normalizacije:
  1. normalizacija dekompozicijom,
  2. normalizacija sintezom.

- *Normalizacija dekompozicijom* započinje od proizvolje nenormalizovane relacione šeme i izvodi se u koracima.
  - Svakim korakom normalizacije relaciona šema se prevodi u višu normalnu formu, tako da se polazni skup atributa deli u dva skupa i od svakog formira posebna relaciona šema.
  - Svaki korak normalizacije mora biti reverzibilan.

- *Normalizacija sintezom* polazi od skupa atributa i od skupa zavisnosti zadatih na tom skupu atributa. Postupak se ne izvodi u koracima već se direktno formiraju relacione šeme koje ispunjavaju uslove zahtevane normalne forme.

NORMALIZACIJA DEKOMPOZICIJOM

- Određivanje kandidata za ključ
  - Sporedni (neključni) atributi su svi oni koji ne pripadaju bilo kom kandidatu za ključ.
- Ukloniti sve redudantne f-ne zavisnosti, tj. one koje se mogu izvesti iz drugih.
  - Odrediti nereduantno pokrivanje skupa FZ.
- Dekompozicija prema f-im zavisnostima od interesa.
- Kada je dekompozicija dobra?
  - Svi atributi iz početne relacije se moraju pojaviti i u novim relacijama.
  - Dekompozicija je dobra samo u slučaju da se ne gube f-ne zavisnosti.
    - Sve FZ polaznog skupa moraju da budu očuvane (direktno ili mogućim izvodjenjem iz skupa relacija dobijenih dekompozicijom)
  - Ako u novodobijenim projekcijama nastalim razbijanjem osnovne relacije postoji zajednički atribut, on mora da bude ključ u bar jednoj od novodobijenih relacija.

ODREĐIVANJE KK

- Ukloniti trivijalne funkcijske zavisnosti iz F
- Rasporediti atribute u četiri grupe:

  i. Elementi koji se ne nalaze ni u jednoj funkcijskoj zavisnosti u F

  ii. Elementi koji se nalaze isključivo sa leve strane funkcijskih zavisnosti u F

  iii. Elementi koji se nalaze isključivo sa desne strane funkcijskih zavisnosti u F

  iv. Elementi koji se nalaze sa obe strane funkcijskih zavisnosti u F
- Izvršiti uniju grupa i i ii i odrediti zatvarač skupa atributa
- A) Ako se u koraku III dobio KK to je jedini KK.

  B) Ako se u koraku III nije dobio KK, onda se kombinuje unija grupa i i ii sa atributima grupe iv. Kombinacije se vrše iterativno po broju atributa.

ODREĐIVANJE KK

R(A, B, C, D, E, F) **F={AB→C, C→B, CD→AD, BD→AE}**.

- Ukloniti trivijalne funkcionalne zavisnosti iz F

  F={AB→C, C→B, CD→A, BD→AE}

- Rasporediti atribute u četiri grupe:

  i. Elementi koji se ne nalaze ni u jednoj f-oj zavisnosti u F

  F

  ii. Elementi koji se nalaze isključivo sa leve strane f-ne zavisnosti u F

  D

  iii. Elementi koji se nalaze isključivo sa desne strane f-ne zavisnosti u F

  E

  iv. Elementi koji se nalaze sa obe strane f-ne zavisnosti u F

  A, B. C

- Izvršiti uniju grupa i i ii i odrediti zatvarač skupa atributa

  (F, D)+→ F, D

- Kombinovanje unija grupa i i ii sa atributima grupe iv.

  (F, D, A)+→ F, D, A (nije KK)

  (F, D, B)+→ F, D, B, A, E, C (KK)

  (F, D, C)+→ F, D, C, B, A, E (KK)

SVOĐENJE NA 2NF POSTUPKOM DEKOMPOZICIJE

R(<u>A,B</u>,C,D, E)
Neka već znamo KK.
A → C
B → D

Kako teče dekompozicija:

1. A,B → C je parcijalna zavisnost i zato
   R(A,B,D) KK(A,B) i B → D
   R1(A,C) A → C
2. A,B → D je parcijalna zavisnost i zato
   R(A,B)
   R2(B,D)
3. R3 = R

Konačno R1(A,C) R2(B,D) R3(A,B)

Algoritam 2NF normalizacije

Ulaz: relacija R sa atributima $\{A_1, \ldots, A_n\}$ i skupom FZ F;

Izlaz: skup $R_1, \ldots, R_i$ $(i \geq 1)$ projekcija relacije R, koje su sve u 2NF i za koje važi: ako je $i = 1$, onda je $R_1 \equiv R$; ako je $i > 1$, onda je $R = R_1 * R_2 * \cdots * R_i$;

BEGIN
  $i := 1$;
  WHILE relacija R nije u 2NF DO
    BEGIN
      uočiti parcijalnu FZ $X \rightarrow A$, $(X \subseteq Atr(R), A \in Atr(R))$
        sporednog atributa A od ključa X;
      neka je $X = X'X''$, gde je $X' \rightarrow A$ potpuna FZ;
      neka je $Z = Atr(R) \setminus (X \cup \{A\})$;
      zameniti relaciju R njenim projekcijama $R[XZ]$, $R[X'A]$;
      $R_i := R[X'A]$;
      $i := i + 1$;
      $R := R[XZ]$
    END;
  $R_i := R$
END.

SVOĐENJE NA 3NF POSTUPKOM DEKOMPOZICIJE

R(A,B,C,D)
B → C
C → D

Neka već znamo KK. Kako teče dekompozicija:

1. A → C,D je tranzitivna zavisnost, zbog
   B → C,D i zato
   R(A,B)
   R1(B,C,D)   B → C,   C → D
2. A,B → D je parcijalna zavisnost i zato
   R(A,B)
   R2(B,D)
3. R3 = R

Konačno R1(A,C)   R2(B,D)   R3(A,B)

Algoritam 3NF normalizacije

Ulaz: relacija R sa atributima $\{A_1, \ldots, A_n\}$ i skupom FZ F;

Izlaz: skup $R_1, \ldots, R_m$ $(m \geq 1)$ projekcija relacije R, koje su sve u 3NF i za koje važi: ako je $m = 1$, onda je $R_1 \equiv R$; ako je $m > 1$, onda je $R = R_1 * R_2 * \cdots * R_m$;

BEGIN
  $i := 1$;
  WHILE relacija R nije u 3NF, DO
    BEGIN
      uočiti tranzitivnu zavisnost $X \to Y, Y \to A, Y \not\to X$ sporednog atributa A od ključa X, za koju je $Y \to A$ potpuna funkcionalna zavisnost;
      neka je $Z = Atr(R) \setminus (X \cup \{A\})$;
      zameniti relaciju R projekcijama $R[YA]$, $R[XZ]$;
      $R_i := R[YA]$;
      $i := i + 1$;
      $R := R[XZ]$
    END;
  $R_i := R$
END.

BOYCE-CODD-OVA NORMALNA FORMA (BCNF)

- Kodova originalna definicija 3NF nije uzimala u obzir slučajeve kada
  - relacija ima više od jednog kandidata za ključ
  - složeni kandidati za ključeve se preklapaju
- Ovi slučajevi su obuhvaćeni Bojs-Kodovom normalnom formom

BOYCE-CODD-OVA NORMALNA FORMA (BCNF)

- Relacija sa tzv "preklapajucim" kandidatima za ključ (dva ili više složenih kandidata za kljuc koji imaju barem jedan zajednicki atribut).

  PRIJAVA(BI, ŠPRED, NAZPRED, OCENA)

  BI, ŠPRED ---> NAZPRED
  BI, ŠPRED ---> OCENA
  ŠPRED ---> NAZPRED
  NAZPRED ---> ŠPRED

  KK(BI, ŠPRED) i KK(BI, NAZPRED)

- Determinanta relacije R je bilo koji atribut, prost ili složen, od koga neki drugi atribut u relaciji potpuno funkcionalno zavisi.
- Relacija R je u Boyce-Codd-ovoj normalnoj formi (BCNF) ako i samo ako su sve determinante u relaciji i kandidati za ključ.

BOYCE-CODD-OVA NORMALNA FORMA (BCNF)

PRIJAVA(BI, ŠPRED, NAZPRED, OCENA)

BI, ŠPRED ---> NAZPRED, OCENA	(D) (KK)
BI, NAZPRED ---> ŠPRED, OCENA	(D) (KK)
ŠPRED ---> NAZPRED	(D)
NAZPRED ---> ŠPRED	(D)

KK(BI, ŠPRED) i KK(BI, NAZPRED)

- Dekompozicijom, pri kojoj se iz relacije “izvlace” projekcije sa onim determinantama koje nisu kandidati za kljuc, relacija se svodi na BCNF.

  PRIJAVA1(BI, ŠPRED, OCENA)
  PREDMET(ŠPRED, NAZPRED)

- Svaka relacija koja je u BCNF je sigurno i u 2NF i 3NF. Obrnuto ne važi.

DEKOMPOZICIJA NA ZAVISNE I NEZAVISNE PROJEKCIJE

- Projekcije R1 i R2 relacije R su nezavisne ako i samo ako važi sledeće:

    1. Svaka funkcionalna zavisnost u R se može logicki dedukovati iz funkcionalnih zavisnosti u R1 i R2 i

    2. Zajednicki atribut relacija R1 i R2 je kandidat za ključ barem u jednoj od njih.

VIŠENAČNE ZAVISNOSTI I ČETVRTA NORMALNA FORMA

PREDMET	NASTAVNIK	KNJIGA
INF-SIST	BRANKO	MARTIN
	KRCA	DATE
	BAJA	
SIST-ANAL	VLADAN	DEMARCO
		SARSON

Jedan predmet predaje više nastavnika. Za jedan predmet se koristi više knjiga. Ne postoji nikakva veza izmedu nastavnika i knjiga.

- Normalizovana

PREDMET	NASTAVNIK	KNJIGA
INF-SIST	BRANKO	MARTIN
INF-SIST	BRANKO	DATE
INF-SIST	KRCA	MARTIN
INF-SIST	KRCA	DATE
INF-SIST	BAJA	MARTIN
INF-SIST	BAJA	DATE
SIST-ANAL	VLADAN	DEMARCO
SIST-ANAL	VLADAN	SARSON

VIŠENAČNE ZAVISNOSTI I ČETVRTA NORMALNA FORMA

RASP(PREDMET, NASTAVNIK) i UDŽBENIK(PREDMET, KNJIGA)

RASP	**PREDMET**	**NASTAVNIK**
	INF-SIST	BRANKO
	INF-SIST	KRCA
	INF-SIST	BAJA
	SIST-ANAL	VLADAN

UDŽBENIK	**PREDMET**	**KNJIGA**
	INF-SIST	MARTIN
	INF-SIST	DATE
	SIST-ANAL	DEMARCO
	SIST-ANAL	SARSON

- Dekompozicija je bez gubljenja informacija.
- Veze koje postoje izmedu atributa ove relacije nazivaju se višeznacnim vezama.

VIŠENAČNE ZAVISNOSTI I ČETVRTA NORMALNA FORMA

- U relaciji R(A, B, C) postoji višeznacna zavisnost A ->-> B ako za datu vrednost A, postoji skup od nula, jedne ili više vrednosti B, a taj skup vrednosti ni na koji nacin ne zavisi od vrednosti atributa C. Atributi A, B i C mogu biti složeni.
- U relaciji R(A,B,C) postoji višeznacna zavisnost A ->-> B ako i samo ako kad god u njoj postoje n-torke

  <a,b,c> i <a,b',c'>,

  postoje takode i n-torke

  <a,b,c'> i <a,b',c>.

  Atributi A, B i C mogu biti složeni.
- U prethodnom primeru važile su

  PREDMET -->--> NASTAVNIK i PREDMET -->--> KNJIGA.

VIŠENAČNE ZAVISNOSTI I ČETVRTA NORMALNA FORMA

- Relacija R je u Četvrtoj normalnoj formi (4NF) ako i samo ako kad god postoji višeznacna funkcionalna zavisnost, na primer A ->-> B, tada svi atributi relacije moraju takođe biti funkcionalno zavisni od A.
- Relacija R je u 4NF ako u njoj nisu date dve (ili više) nezavisne višeznačne činjenice.

ZAVISNOSTI SPAJANJA I PETA NORMALNA FORMA (5NF)

- U relaciji R(X,Y, ..., Z) postoji zavisnost spajanja ako i samo ako relacija R rezultuje iz prirodnog spajanja njenih projekcija po X,Y, ..., Z, gde su X,Y, ..., Z podskupovi atributa relacije R.
- Relacija R je u Petoj normalnoj formi ako i samo ako se svaka zavisnost spajanja može pripisati kandidatu za ključ.