NORMALIZACIJA

NORMALIZACIJA

- Normalizacija je postupak logičkog projektovanja baze podataka kojim se odstranjuju anomalije u njenom održavanju.
- Redudansa (ponavljanje, dupliranje) podataka je jedan od glavnih uzroka pojave anomalija.
- Određena struktura baze podataka se smatra dobrom ako joj je logička redudansa minimalna.

NENORMALIZOVANA RELACIJA I ANOMALIJE ODRŽAVANJA (2)

Nenormalizovana relacija je relacija koja poseduje neke vrednosti atributa koje nisu "atomske", odnosno relacija koja poseduje "grupe sa ponavljanjem".

STUDENT

BI	IME	SEM	ŠSMER	IMERUK	POLOZIO (ŠPRED-NAZPRED-OCENA)		
21	ZORAN	5	01	BATA	{121,	MATEMAT,	7}
					{323,	BAZEPOD,	8}
					{056,	SOCIOL,	8}
77	ANA	7	01	BATA	{056,	MARKSIZ,	10}
					{121,	MATEMAT,	5}
36	PERA	4	02	MIKA	{323,	BAZEPOD,	8}
					{456,	ELEKTRON,	9}
					{442,	FIZIKA,	6}
					{056,	SOCIOL,	8}

Jedna n-torka ove nenormalizovane relacije bi se mogla zapisati ovako:
(21, ZORAN, 5, 01, BATA, { (121,MATEMAT,7), (323,BAZEPOD,8), (056, SOCIOL,8) })

NENORMALIZOVANA RELACIJA I ANOMALIJE ODRŽAVANJA (3)

- Anomalije u dodavanju:
  Ako je u novom nastavnom planu definisan novi predmet, ne mogu se ubaciti podaci o tom predmetu dok ga neki student ne položi.
  Ili, ako se otvori neki novi smer, ne mogu se ubaciti podaci o tom smeru dok ga neki student ne upiše.
- Anomalije u brisanju:
  Ako je jedan predmet (FIZIKA) položio samo jedan student (PERA) i ako se on ispiše sa fakulteta, odnosno izbaci odgovarajuca n-torka, gube se i sve informacije o tom predmetu.
  Ako je taj student bio i jedini student na nekom smeru, gube se i sve informacije o tom smeru.
- Anomalije u ažuriranju:
  Ako se promeni naziv nekog predmeta ili rukovodioc nekog smera, to se mora uciniti na onoliko mesta koliko je studenata položilo taj predmet, odnosno koliko je studenata upisano na dati smer.
- Izveštavanje za datu strukturu:
  - Data struktura relacije je veoma pogodna za izveštaj na zahtev za izdavanje uverenja o položenim ispitima.
  - "Prikaži listu predmeta, imena svih studenata koji su ga položili i prosecnu ocenu na predmetu", za datu strukturu relacije zahtevao bi znatno složeniji program.

ČEMU SLUŽI NORMALIZACIJA

- Postupkom normalizacije logicka struktura baze podataka se dovodi u takav oblik (ili, drugim recima, relacije se dovode u normalne forme) u kome se izbegavaju anomalije u održavanju i problemi u izveštavanju.

PRVA NORMALNA FORMA (1NF)

- Relacija R je u Prvoj normalnoj formi (1NF) ako su sve vrednosti njenih atributa atomske.
- Termin normalizovana relacija se koristi za relacije koje su u 1NF.
- Svi upitni jezici zasnovani na relacionoj algebr zahtevaju normalizovane relacije.

PRVA NORMALNA FORMA (1NF)

- **Relacija R je u Prvoj normalnoj formi (1NF) ako su sve vrednosti njenih atributa atomske**.
- Termin normalizovana relacija se koristi za relacije koje su u 1NF.
- Svi upitni jezici zasnovani na relacionoj algebri zahtevaju (horizontalno) normalizovane relacije.

STUDENT

BI	IME	SEM	ŠSMER	IMERUK	ŠPRED	NAZPRED	OCENA
21	ZORAN	5	01	BATA	121	MATEMAT	7
21	ZORAN	5	01	BATA	323	BAZEPOD	8
21	ZORAN	5	01	BATA	056	SOCIOL	8
77	ANA	7	01	BATA	056	MARKSIZ	10
77	ANA	7	01	BATA	121	MATEMAT	5
36	PERA	4	02	MIKA	323	BAZEPOD	8
36	PERA	4	02	MIKA	456	ELEKTRON	9
36	PERA	4	02	MIKA	442	FIZIKA	6
36	PERA	4	02	MIKA	056	SOCIOL	8

FUNKCIONALNE (FUNKCIJSKE) ZAVISNOSTI

Data je relacija R sa atributima X i Y, moguce složenim. Atribut Y je funkcionalno zavisan od atributa X (ili X funkcionalno odreduje Y),

R.X -> R.Y,

ako i samo ako svakoj vrednosti X odgovara jedna i samo jedna vrednost Y.

Primer.

STUDENT1 (BI, IME, SEM, ŠSMER, IMERUK) (1)

BI -> IME

BI -> SEM

BI -> ŠSMER

BI -> IMERUK

Atribut Y relacije R je funkcionalno zavisan od atributa X relacije R ako i samo ako kad god dve n-torke relacije R imaju istu x-vrednost one moraju imati istu i y-vrednost.

ZADOVOLJENOST F-NE ZAVISNOSTI

Ako u relaciji R važi funkcionalna zavisnost $X \rightarrow Y$, 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 $X \rightarrow Y$ 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 $X \rightarrow Y$ ako svako njeno stanje r zadovoljava tu FZ.

PITANJE

Da li dato stanje relacije zadovoljava sledeću f-nu zavisnost:

a) $\{C\} \rightarrow \{A\}$

b) $\{A, D\} \rightarrow \{C\}$

c) $\{B\} \rightarrow \{E\}$

A	B	C	D	E
1	1	1	1	1
1	2	2	2	1
2	1	3	3	1
2	1	4	3	1
3	2	5	1	1

PITANJE

Da li dato stanje relacije zadovoljava sledeću f-nu zavisnost:

a) $\{C\} \rightarrow \{A\}$

Važi. Postojanje 3 -> 2, 4 -> 2 nije u suprotnosti sa zahtevom definicije

b) $\{A, D\} \rightarrow \{C\}$

c) $\{B\} \rightarrow \{E\}$

A	B	C	D	E
1	1	1	1	1
1	2	2	2	1
2	1	3	3	1
2	1	4	3	1
3	2	5	1	1

PITANJE

Da li dato stanje relacije zadovoljava sledeću f-nu zavisnost:

a) $\{C\} \rightarrow \{A\}$

Važi. Postojanje 3 -> 2, 4 -> 2 nije u suprotnosti sa zahtevom definicije

b) $\{A, D\} \rightarrow \{C\}$

Ne važi jer postoji (2,3) -> 3, (2,3) -> 4

c) $\{B\} \rightarrow \{E\}$

A	B	C	D	E
1	1	1	1	1
1	2	2	2	1
2	1	**3**	**3**	1
2	1	**4**	**3**	1
3	2	5	1	1

PITANJE

Da li dato stanje relacije zadovoljava sledeću f-nu zavisnost:

a) $\{C\} \rightarrow \{A\}$

Važi. Postojanje 3 -> 2, 4 -> 2 nije u suprotnosti sa zahtevom definicije

b) $\{A, D\} \rightarrow \{C\}$

Ne važi jer postoji (2,3) -> 3, (2,3) -> 4

c) $\{B\} \rightarrow \{E\}$

Važi.

A	B	C	D	E
1	1	1	1	1
1	2	2	2	1
2	1	3	3	1
2	1	4	3	1
3	2	5	1	1

FUNKCIONALNE (FUNKCIJSKE) ZAVISNOSTI

X, Y, Z - podskupovi skupa atributa relacije R

- $X \rightarrow Y$ 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 $Z \rightarrow Y$

POTPUNE FUNKCIONALNE ZAVISNOSTI

Atribut Y relacije R je potpuno funkcionalno zavisan od kolekcije atributa X relacije R ako je funkcionalno zavisan od kolekcije atributa X, a nije funkcionalno zavisan ni od jednog pravog podskupa kolekcije atributa X.

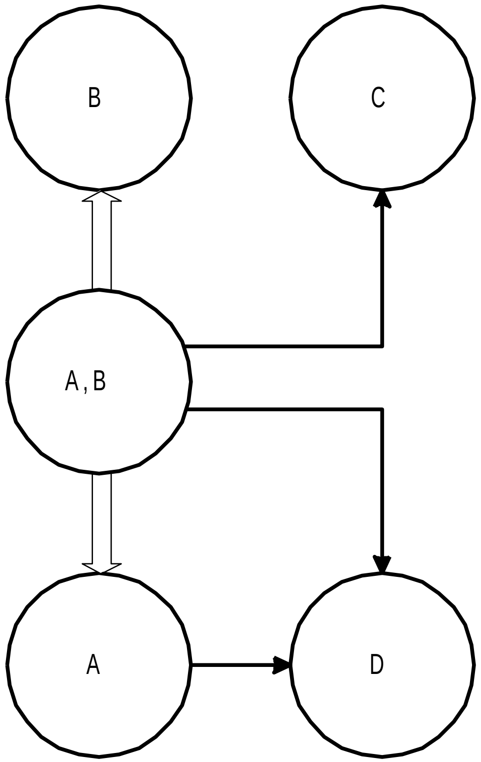
Primere potpune i nepotpune funkcionalne zavisnosti možemo pokazati na relaciji PRIJAVA:

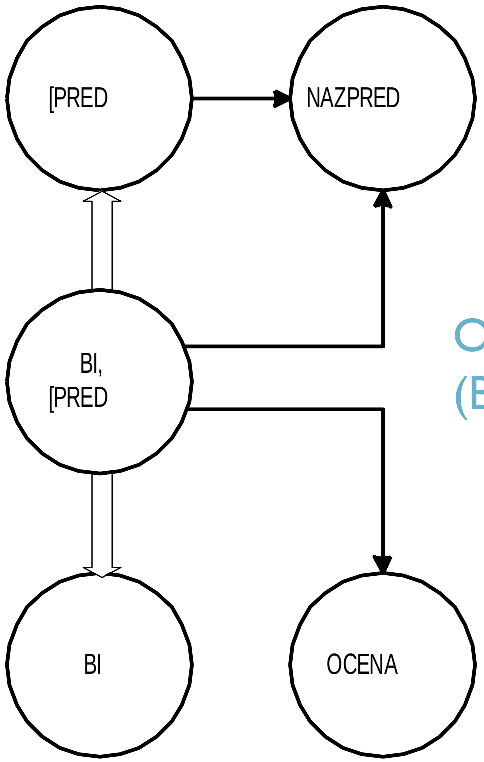
(a) {BI, ŠPRED} -> {OCENA}
{BI} -/-> {OCENA}
{ŠPRED} -/-> {OCENA}
atribut OCENA je potpuno funkcionalno zavisnan od {BI, ŠPRED}.

(b) {BI, ŠPRED} -> {NAZPRED}
{BI} -/-> {NAZPRED}
{ŠPRED} ---> {NAZPRED}
atribut NAZPRED je nepotpuno funkcionalno zavisan od složenog atributa BI, ŠPRED, jer je funkcionalno zavisan i od njega i od jednog njegovog dela, od ŠPRED.

POTPUNE FUNKCIONALNE ZAVISNOSTI

C je potpuno funkcionalno zavisno od složenog atributa (A,B), a D nije.



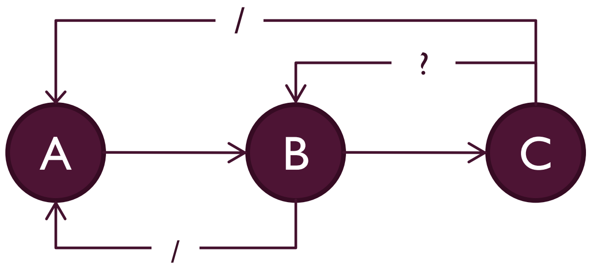


OCENA je potpuno zavisna od (BI, ŠPRED), a NAZPRED nije.

sa ⇒ su ozna~ene trivijalne funkcije

TRANZITIVNE FUNKCIONALNE ZAVISNOSTI

Atribut C je tranzitivno funkcionalno zavisan od atributa A ako je funkcionalno zavisan od A i ako je funkcionalno zavisan od nekog atributa B koji je i sam funkcionalno zavisan od A.

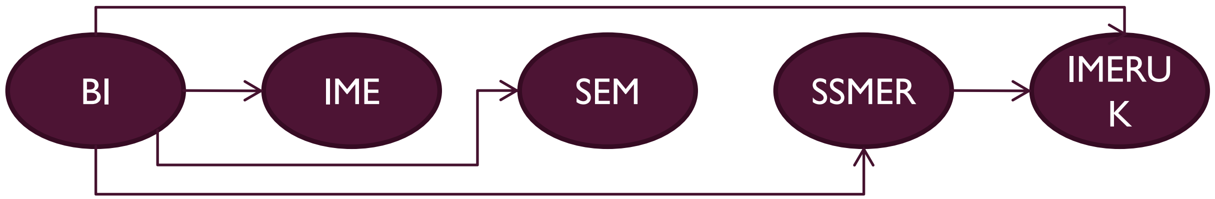


Tranzitivna funkcionalna zavisnost je redundantna, pa je nije neophodno pamtiti u bazi podataka.

f1: BI ---> ŠSMER

f2: ŠSMER ---> IMERUK

f3 = f1 ◦ f2 : BI ---> IMERUK



DEFINICIJA KLJUČA I NADKLJUČA

- Atribut X moguce složeni, je nadkljuc neke relacije R ako i samo ako funkcionalno odreduje sve ostale atribute relacije R.
- Atribut X, moguce složeni, je kljuc relacije R ako je nadkljuc relacije R, a nijedan njegov pravi podskup nema tu osobinu.

NORMALNE FORME

- Relacija R je u **Drugoj normalnoj formi (2NF)** ako i samo ako je u **1NF** i svi njeni neključni atributi **potpuno funkcionalno** zavise od primarnog ključa.

NORMALNE FORME

- Relacija R je u **Trecoj normalnoj formi (3NF)** ako i samo ako je u **2NF** i ako svi njeni nekljucni atributi **netranzitivno funkcionalno** zavise od ključa.

NORMALNE FORME

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

- Relacija sa tzv "preklapajucim" kandidatima za ključ (dva ili više složenih KK koji imaju barem jedan zajednicki atribut) ne mora biti u koliziji sa 2NF i 3NF, ali nije u 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)

NORMALNE FORME

- IVNF i VNF na sledećem času.

VERTIKALNA NORMALIZACIJA

VERTIKALNA NORMALIZACIJA

- Cilj: formiranje relacionih šema koje ispunjavaju uslove zahtevane normalne forme.

$R(A, B, C, D)$

$A, B \rightarrow A, B, C, D$

$A \rightarrow C$

$\longrightarrow$

$R_1(A, B, D)$

$A, B \rightarrow A, B, D$

$R_2(A, C)$

$A \rightarrow C$

POSTUPCI VERTIKALNE NORMALIZACIJE

- Postupci vertikalne normalizacije:
  1. **normalizacija dekompozicijom,**
  2. normalizacija sintezom.

- *Normalizacija dekompozicijom* započinje od proizvoljne 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 (dekompozicija bez gubljenja informacija).

- *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.

DEKOMPOZICIJA RELACIJA BEZ GUBLJENJA INFORMACIJA

- Relacija R se dekomponuje u svoje projekcije bez gubljenja informacija ako prirodno spajanje tako dobijenih projekcija dovodi do polazne relacije.

PRIJAVA

BI	ŠPRED	NAZPRED	OCENA
21	121	MATEMAT	7
21	323	BAZEPOD	8
21	056	SOCIOL	8
77	056	SOCIOL	10
77	121	MATEMAT	5
36	323	BAZEPOD	8
36	456	ELEKTRON	9
36	442	FIZIKA	6
36	056	SOCIOL	8

R1

BI	ŠPRED	OCENA
21	121	7
21	323	8
21	056	8
77	056	10
77	121	5
36	323	8
36	456	9
36	442	6
36	056	8

R2

NAZPRED	OCENA
MATEMAT	7
BAZEPOD	8
SOCIOL	8
SOCIOL	10
MATEMAT	5
ELEKTRON	9
FIZIKA	6

DEKOMPOZICIJA RELACIJA BEZ GUBLJENJA INFORMACIJA

R3 = R1 [OCENA * OCENA] R2

R3	BI	ŠPRED	NAZPRED	OCENA
	21	121	MATEMAT	7
	21	323	BAZEPOD	8
	21	056	SOCIOL	6
	77	056	SOCIOL	10
	77	121	MATEMAT	5
	36	323	BAZEPOD	9
	36	456	ELEKTRON	9
	36	442	FIZIKA	6
	36	056	SOCIOL	8
	------------	------------	------------	------------
	21	323	SOCIOL	8
	36	056	BAZEPOD	8
	21	056	FIZIKA	6
	36	442	SOCIOL	6
	36	323	ELEKTRON	9
	36	465	BAZEPOD	9

DEKOMPOZICIJA RELACIJA BEZ GUBLJENJA INFORMACIJA

- Heath-ova teorema

  Relacija R(A,B,C), gde su A, B i C podskupovi atributa, u kojoj važi R.A --> R.B može se uvek dekomponovati u svoje projekcije R1(A,B) i R2(A,C) bez gubljenja informacija.

- Za navedeni primer relacije PRIJAVA dekompozicija bez gubljenja informacija bi bila dekompozicija u projekcije
  R1(BI, ŠPRED, OCENA) i R2(ŠPRED, NAZPRED)

NORMALIZACIJA DEKOMPOZICIJOM

- 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.

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.

Za ovo nam je potrebna teorija zavisnosti!

VERTIKALNA NORMALIZACIJA

ARMSTRONGOVE AKSIOME IZVOĐENJA

- Aksiome izvođenja su pravila koja utvrđuju važenje nekih funkcionalnih zavisnosti u relaciji R, na osnovu važenja nekih drugih funkcionalnih zavisnosti u R.
- Mogu se izvesti na osnovu osobina f-nih zavisnoti, pa u svojoj suštini nisu prave aksiome.
- Armstrongove aksiome (u modifikovanom obliku) su:

  - Aksioma refleksivnosti:
    Funkcionalna zavisnost **X->X** uvek važi.

  - Aksioma proširenja:
    Ako u relaciji R važi **X->Y** i ako je Z podskup skupa Atr(R), onda važi i **X,Z -> Y** .

  - Aksioma aditivnosti:
    Ako u relaciji R važi **X->Y** i **X->Z**, onda važi i **X -> Y,Z** .

  - Aksioma projektivnosti:
    Ako u relaciji R važi **X->Y,Z**, onda važi i **X -> Y** i **X -> Z**.

ARMSTRONGOVE AKSIOME IZVOĐENJA

- Aksioma tranzitivnosti:
  Ako u relaciji R važi **X->Y** i **Y->Z**, onda važi i **X -> Z** .

- Aksioma pseudotranzitivnosti:
  Ako u relaciji R važi **X->Y** i **Y,Z->W**, onda važi i **X, Z-> W** .

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 →B, A →C}

F^+ = { A → A, A → B, A → C, A,B → A, A,B → B,
A,B → C A,B → A, B A,B → B,C A,B → A,C,
A,B → A,B,C, A, C → ..., A,B,C → 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

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

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 $X \rightarrow Y$ 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.

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.

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)