

Projektovanje baza podataka

Logičko modeliranje – Prečišćavanje sheme

3. jun 2026.

Sadržaj

- 1 Višeznačne zavisnosti i zavisnosti spajanja
- 2 Više normalne forme
- 3 Postupci i algoritmi

Napomena

Slajdovi su nastali obradom slajdova prof. dr Saše Malkova za predmet Projektovanje baza podataka

Normalne forme

- **Normalne forme** su specifični oblici relacija koji zadovoljavaju određena pravila u vezi funkcionalnih zavisnosti
 - koji zato **garantuju** da u relaciji neće biti redundantnosti određenog tipa

Normalne forme I

- U svom prvom radu o relacionom modelu (1970), Kod je definisao samo
 - 1. normalnu formu (1NF)
- Godinu dana kasnije (1971) je definisao
 - 2. normalnu formu (2NF)
 - 3. normalnu formu (3NF)
- Zajedno sa Rejmondom Bojsom je definisao (1974)
 - Bojs-Kodovu normalnu formu (BCNF)
- Kasnije su formulisane još i
 - 4. normalna forma (4NF) (1977)
 - 5. normalna forma (5NF) (1979)
 - i druge više normalne forme

Sadržaj

- 1 Višeznačne zavisnosti i zavisnosti spajanja
- 2 Više normalne forme
- 3 Postupci i algoritmi

Ograničenja funkcionalnih zavisnosti

- Kao što njihovo ime kaže, FZ opisuju zavisnosti koje imaju oblik funkcije (preslikavanja):
 - za datu vrednost jednog (skupa) atributa određuju tačno jednu vrednost drugog (skupa) atributa
- To nije dovoljno za opisivanje svih zavisnosti u nekom posmatranom domenu ili u BP koja ga modelira

Ograničenja funkcionalnih zavisnosti (3)

- Formalna potvrda da FZ nisu dovoljne je činjenica da:
 - FZ jesu dovoljan uslov za potpunu dekompoziciju
 - FZ **nisu** neophodan uslov za potpunu dekompoziciju
- Na primer, neka relacija ilustruje podatke o pogledanim filmovima:

Gledalac	Naslov	Glumac
Goran	Balkanski špijun	Danilo Bata Stojković
Dragana	Balkanski špijun	Danilo Bata Stojković
Goran	Balkanski špijun	Bora Todorović
Dragana	Balkanski špijun	Bora Todorović
Goran	Profesionalac	Bora Todorović
Goran	Profesionalac	Branislav Lečić

Tabela: Primer relacije Film(Gledalac, Naslov, Glumac)

- Ne postoji FZ, ali je moguća potpuna dekompozicija na:
 - {Gledalac, Naslov}, {Naslov, Glumac}

Višeznačne zavisnosti

- **Višeznačne zavisnosti** se uvode zato što FZ nisu dovoljne da opišu sve redundantnosti
 - a time ni sve uslove za potpunu dekompoziciju
- Ideja je da svaka vrednost jednog skupa atributa X određuje **skup vrednosti** drugog skupa atributa Y
 - funkcionalne zavisnosti predstavljaju specijalan slučaj višeznačnih zavisnosti, kada je svaki zavisani skup jednočlan

Višeznačne zavisnosti (def)

- Definicija

- Između skupa atributa X i zavisnog skupa atributa Y relacije R (gde su $X, Y \subseteq \text{Attr}(R)$ i $Z = \text{Attr}(R) \setminus XY$) postoji **višeznačna zavisnost**, u oznaci $X \twoheadrightarrow Y$, akko za svake dve torke t_1 i t_2 iz R takve da je $t_1[X] = t_2[X]$ postoji torka t_3 takva da je $t_3[X] = t_1[X] = t_2[X]$ i $t_3[Y] = t_1[Y]$ i $t_3[Z] = t_2[Z]$.
- koristićemo i oznaku $X \twoheadrightarrow Y$

Višeznačne zavisnosti (2)

- Neformalno rečeno, za svaku izabranu vrednost atributa iz X , vrednosti iz Y su nezavisne od vrednosti iz Z , tj. zavise samo od X .
- Ako postoje torke (x, y_1, z_1) i (x, y_2, z_2) , onda postoji i torka (x, y_1, z_2)
 - posledica je da postoji i torka (x, y_2, z_1)
- T.j. ako su X , Y i Z disjunktni skupovi atributa koji pokrivaju $\text{Attr}(R)$, onda važi $\forall Z \ X \twoheadrightarrow Y$ akko za svaku fiksiranu vrednost X_c restrikcija po uslovu $X=X_c \ \&\& \ Z=Z_c$ ima istu vrednost, bez obzira na izabrano Z_c .

Višeznačne zavisnosti (primer)

- Na prethodnom primeru možemo da uočimo višeznačne zavisnosti:
 - $\{\text{Naslov}\} \twoheadrightarrow \{\text{Glumac}\}$
 - $\{\text{Naslov}\} \twoheadrightarrow \{\text{Gledalac}\}$

Gledalac	Naslov	Glumac
Goran	Balkanski špijun	Danilo Bata Stojković
Dragana	Balkanski špijun	Danilo Bata Stojković
Goran	Balkanski špijun	Bora Todorović
Dragana	Balkanski špijun	Bora Todorović
Goran	Profesionalac	Bora Todorović
Goran	Profesionalac	Branislav Lečić

Tabela: Primer relacije Film(Gledalac, Naslov, Glumac)

VZ i dekompozicija

- Teorema:
 - Neka su X, Y, Z neprazni skupovi atributa relacije R , i pri tome je $Z = \text{Attr}(R) \setminus XY$.
 Onda na relaciji R važi jednakost $R = R[XY] * R[XZ]$ akko važi $VZ X \twoheadrightarrow Y$

VZ i dekompozicija (2)

- Veoma je važno što postoji ekvivalencija (akko) između VZ i potpune dekompozicije jedne relacije na dve relacije
- Zbog toga je ustanovljavanje VZ dovoljan i neophodan uslov za potpunu dekompoziciju jedne relacije na dve manje relacije
- Ako je X neprazan skup atributa, onda kažemo da su $VZ\ X \twoheadrightarrow \emptyset$ i $X \twoheadrightarrow Attr(R) \setminus X$ **trivijalne**

VZ nisu dovoljne

- U nekim slučajevima jedna relacija ne može da se potpuno dekomponuje na dve relacije ali može na tri
- tj. VZ ipak ne predstavljaju najopštiji oblik zavisnosti

Zavisnosti spajanja

- Neka su $X_1, X_2, \dots, X_n$ skupovi atributa relacije R , čija je unija upravo $\text{Attr}(R)$. U relaciji R važi **zavisnost spajanja**, u oznaci $\ast\{X_1, X_2, \dots, X_n\}$ akko je: $R = R[X_1] \ast R[X_2] \ast \dots \ast R[X_n]$
- Zavisnosti spajanja su definisane tako da predstavljaju **najopštiji mogući oblik zavisnosti** – *neophodan i dovoljan za potpunu dekompoziciju*
 - zavisnosti spajanja sa dva skupa atributa su ekvivalentne višeznačnim zavisnostima
- Zavisnost spajanja $\ast\{X_1, X_2, \dots, X_n\}$ je **trivijalna** akko za neki od skupova atributa X_i važi $X_i = \text{Attr}(R)$.

Reducibilna zavisnost spajanja

- Zavisnost spajanja je reducibilna akko postoji pravi podskup njenih komponenti, koji i dalje određuje (redukovanu) zavisnost spajanja
 - ekvivalentno:
 - Zavisnost spajanja $\ast\{X_1, X_2, \dots, X_n\}$ je reducibilna akko neka komponenta X_i može da se isključi, a da i dalje imamo zavisnost spajanja $\ast\{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}$.
- Na primer, ako je X_i pravi podskup neke druge komponente, onda je ZS reducibilna
- Zanimaju nas samo nereducibilne zavisnosti spajanja

Sadržaj

- 1 Višeznačne zavisnosti i zavisnosti spajanja
- 2 Više normalne forme
- 3 Postupci i algoritmi

Više normalne forme

- Prvih nekoliko normalnih formi smo definisali na osnovu funkcionalnih zavisnosti:
 - 1. NF
 - čak i bez funkcionalnih zavisnosti
 - 2. NF
 - 3. NF
 - NFEK
 - BKNF

Više normalne forme (2)

- Ostale NF se nazivaju i **više normalne forme**
 - Definišu se na osnovu složenijih oblika zavisnosti
 - 4.NF
 - 5.NF

4. normalna forma

- Relacija R je u 4.NF akko za svaku višeznačnu zavisnost $X \twoheadrightarrow Y$ važi da je X ključ ili natključ.
- U suštini, ako relacija ne zadovoljava 4.NF, to znači da se može definisati kao rezultat spajanja dveju jednostavnijih relacija

4. normalna forma (primer 1)

- Neka imamo relaciju
 - Film(**gledalac**, **naslov**, **glumac**)
- Videli smo da važe zavisnosti
 - { naslov } $\twoheadrightarrow$ { gledalac }
 - { naslov } $\twoheadrightarrow$ { glumac }
 - narušena je 4NF, zato što { naslov } nije (nat)ključ
- onda se Film može normalizovati dekompozicijom na:
 - FilmGledalac(naslov, gledalac)
 - FilmGlumac(naslov, glumac)

4. normalna forma (primer 2)

- Neka imamo relaciju
- Ponuda(prodavnica, proizvod, grad)
- i neka važi pravilo
- $\{ \text{prodavnica} \} \rightarrow \{ \text{proizvod} \}$
- tj. isti skup proizvoda se nudi nezavisno od grada
- tj. u svakom gradu se nude svi proizvodi koje prodavnica nudi
- naravno, onda mora da važi i $\{ \text{prodavnica} \} \rightarrow \{ \text{grad} \}$
- narušena je 4NF, zato što $\{ \text{prodavnica} \}$ nije (nat)ključ
- onda se Ponuda može normalizovati dekompozicijom na:
- PonudaProizvod(prodavnica, proizvod)
- PonudaGrad(prodavnica, grad)

4. normalna forma (diskusija)

- Eventualna nezadovoljenost 4.NF je relativno lako intuitivno uočljiva pažljivim posmatranjem
 - zato se često zaobilazi u osnovnoj nastavi baza podataka
 - kao i višeznačne zavisnosti
 - očekuje se da će problem biti uočen i rešen intuitivno
 - ali to ipak nije slučaj i ovakvi problemi se pojavljuju u produkcionim bazama podataka
- Nezadovoljavanje 4.NF može da dovede do drastičnog narušavanja performansi, zbog značajnog povećavanja broja redova u konkretnim tabelama...
- ...naravno, i do problema sa održavanjem usled redundantnosti

5. normalna forma

- Relacija R je u 5.NF akko je svaka netrivialna zavisnost spajanja logički implicirana ključevima relacije R .
 - Zavisnost spajanja je implicirana ključevima akko svaka komponenta zavisnosti spajanja predstavlja natključ relacije
 - Poznata je i pod imenom *Normalna forma projektivnog spajanja* (PJNF).

5. normalna forma (primer 1)

- Primer relacije u 5.NF može da bude svaka relacija u kojoj su
 - neključni atributi (svaki za sebe) funkcionalno zavisni od ključa
 - npr. Student(indeks, ime, prezime)
 - netrivialna ZS je $*\{ \{ \text{indeks, ime} \} , \{ \text{indeks, prezime} \} \}$

5. normalna forma (primer 2)

- Neka ponovo imamo primer sa prodavnicama:
 - Ponuda(**prodavnica**, **proizvod**, **grad**)
- i neka važi pravilo:
 - ako prodavnica A isporučuje proizvod B i obavlja isporuke u grad C, i pri tome neko isporučuje proizvod B u grad C, onda obavezno i prodavnica A isporučuje proizvod B u grad C
- Relacija jeste u 4.NF – nema višeznačnih zavisnosti
 - ne postoje VZ $\{\text{grad}\} \twoheadrightarrow \{\text{proizvod}\}$ i $\{\text{grad}\} \twoheadrightarrow \{\text{prodavnica}\}$ zbog “ako prodavnica A isporučuje proizvod B i obavlja isporuke u grad C...”
- Relacija nije u 5.NF – postoji dekompozicija (zavisnost spajanja):
 - $*(\{ \text{prodavnica}, \text{proizvod} \}, \{ \text{prodavnica}, \text{grad} \}, \{ \text{proizvod}, \text{grad} \})$
- Problem se rešava navedenom dekompozicijom na tri relacije
- Ne postoji dekompozicija na dve relacije
 - inače bi postojala VZ

Sadržaj

- 1 Višeznačne zavisnosti i zavisnosti spajanja
- 2 Više normalne forme
- 3 Postupci i algoritmi

Normalizacija

- Osnovni cilj prečišćavanje sheme u slučaju relacionog modela je eliminacija redundantnosti iz sheme baze podataka
- Sredstvo kojim se to ostvaruje je **normalizacija logičkog modela**
- Normalizacija logičkog modela baze podataka je transformisanje logičkog modela u ekvivalentan skup relacija koji zadovoljava izabrane normalne forme

Normalizacija (2)

- Obično se kao cilj normalizacije postavlja BKNF ili 4.NF

Normalizacija do BKNF

- Normalizacija logičkog modela baze podataka do BKNF je svođenje na skup relacija koje su sve u Bojs-Kodovoj normalnoj formi
 - Efektivan algoritam
 - 1. Neka je R relacija koja nije u BKNF
 - i neka je X pravi podskup njenih atributa koji nije natključ
 - i A jedan atribut koji zavisi od X (i tako narušava BKNF),
 - onda je potrebno dekompozicijom podeliti R na relacije
 - sa atributima XA i $R \setminus A$
 - ako se tako ne gube neke FZ
 - 2. Ako $R \setminus A$ ili XA nisu u BKNF, dekomponovati ih dalje rekursivnom primenom istog algoritma.

Slabosti normalizacije do BKNF

- U nekim slučajevima **ne postoji** dekompozicija u BKNF koja čuva sve zavisnosti
 - tada smo prinuđeni
 - ili da ostanemo u nižoj NF
 - ili da izgubimo neke zavisnosti
 - ili da promenimo semantiku funkcionalnih zavisnosti

Slabosti BKNF – primer I

- Neka imamo relaciju:
 - Prijavalspita(**student, ispitni rok**, predmet)
 - ključ je (student, ispitni rok)
- i zavisnosti:
 - $\{ \text{student, ispitni rok} \} \rightarrow \{ \text{predmet} \}$
 - tj. student u jednom isp.roku može da prijavi ispit iz najviše jednog predmeta
 - $\{ \text{student, ispitni rok} \}$ je ključ
 - $\{ \text{predmet} \} \rightarrow \{ \text{ispitni rok} \}$
 - tj. svaki predmet se polaže samo u po jednom ispitnom roku
- Tada:
 - relacija Prijavalspita nije u BKNF:
 - ispitni rok zavisi od predmeta, a predmet nije deo ključa
 - ako napravimo dekompoziciju, narušićemo prvu zavisnost...
 - ako promenimo ključ (student, predmet), problem opstaje

Automatizacija

- Neki segmenti posla u vezi sa prečišćavanjem sheme mogu da se automatizuju
 - ustanovljavanje FZ na osnovu uzorka podataka
 - izračunavanje zatvorenja skupa FZ
 - na osnovu početnog skupa “pretpostavljenih” FZ
 - izračunavanje skupa zavisnih atributa
 - pronalaženje ključeva kandidata
 - računa se skup zavisnih atributa za svaki podskup atributa relacije
 - H je natključ akko je $H^+ = \text{Attr}(R)$

Literatura za temu

- Ramakrishnan, Gehrke, Database Management Systems, 2.ed, 2000.
- Gordana Pavlović-Lažetić, Uvod u relacione baze podataka, 2.izd. Matematički fakultet, 1999.
- dostupno onlajn: <http://poincare.matf.bg.ac.rs/~gordana//urbp-2016.htm>