

**Brano Markić**

Docent na Ekonomskom fakultetu u Mostaru

**Ivan Pavlović**

Docent na Ekonomskom fakultetu u Mostaru

## **NORMALIZACIJA PODATAKA KAO TEMELJNI POKAZATELJ VALJANOSTI RELACIJSKOG MODELA PODATAKA**

UDK/UDC 681.3

Pregledni rad

Primljeno/Received: 24. rujna 1996.

### **Sažetak**

*Donositelji odluka na različitim razinama organiziranosti, menadžeri i korisnici, zahtijevaju informacije za donošenje djelotvornih odluka. Pouzdane, relevantne i pravodobne informacije potreban su uvjet odlučivanja. Mi ćemo koncentrirati našu pozornost na organizaciju podataka iz kojih nastaju informacije. Općenito, koncept upravljanja bazama podataka je isti za velike i mikrokompjutore.*

*Razumijevanje sustava baza podataka omogućava njegovu djelotvornu uporabu na radnom mjestu. Ranije se nije postavljao takav zahtjev korisniku. Razvitak moćnih mikrokompjutora i njihovih mreža postavio je zadatak razumijevanja organizacije i strukture baza podataka.*

*Kompetentnost korištenja informacija u informacijskom dobu ima za pretpostavku razumijevanje pojmova datoteka, baza podataka, sustava za upravljanje bazama podataka, modela podataka i drugih pojmova potrebnih za kreiranje i pronalaženje informacija.*

**Ključne riječi:** *model podataka, normalizacija, normalne forme.*

## **1 MODELI PODATAKA**

Razumijevanje kompleksnosti procesa i fenomena u realitetu podrazumijeva prethodno kreiranje modela procesa ili stanja. Modeli su, općenito kazano, sredstva ekonomiziranja kojima se reducira broj dimenzija pojave, problema ali samo onih dimenzija koje ne utječu na njihovo

potpuno razumijevanje. Modeli su sredstva za podršku racionalnom mišljenju. Modeli podataka su važni u svim fazama životnog ciklusa informacijskog sustava. U fazi sustavne analize oni su sredstva opisivanja realnog sustava a istodobno i rezultat opisivanja. U fazi dizajniranja sustava model podataka je temelj oblikovanja baza podataka i njihova predstavljanja odabranim sustavom za upravljanje bazama podataka. U fazi implementacije model podataka je predstavljen u formi komunikacije korisnika s kompjutorom (interfejsa).

Proučavanjem modela podataka došlo se do spoznaje da svaki model podataka čine tri neodvojiva dijela:

- a) struktura podataka,
- b) operacije nad podacima,
- c) ograničenja.

Struktura podataka u modelu podataka opisuje elemente modela, objekte, entitete koji se selektiraju iz realiteta. Strukturu modela sačinjavaju objekti, njihove osobine, veze između objekata i njihovih osobina (M., Varga (1994) "Baze podataka: Konceptualno, logičko i fizičko modeliranje podataka)

Operacije nad podacima su operacije nad strukturom modela kojima se izražava dinamika realnog sustava. Strukturu i ograničenja opisuju stanje realnog sustava (statički ops stanja). Operacije su bitne za dinamiku realnog sustava.

Ograničenja se nekad promatraju kao dio strukture modela podataka zato što razdvajaju dozvoljena od nedozvoljenih stanja realnog sustava (D. Vaskovitch (1995) "Klijent server strategije").

## 2 NORMALIZACIJA PODATAKA

Normalizacija je postupak kojim se eliminira potreba za višestrukim zapisivanjem istih podataka u bazi. Cilj je oblikovati bazu podataka u kojoj se jedan podatak unosi samo jedanput. U radu sa bazom podataka se pojavljuju određene anomalije:

- a) anomalija dodavanja novih podataka u bazu,
- b) anomalija brisanja podataka,
- c) anomalija izmjene sadržaja.

Anomalija dodavanja novih podataka u bazu se javlja onda kada su podaci o jednom entitetu memorirani u bazi kao dio opisa nekog drugog entiteta. Da bi smo bolje razumjeli navedene anomalije prikazat ćemo strukturu baze koja je loše oblikovana. Pretpostavimo da su podaci o nekom studentu strukturirani na sljedeći način:

broj indeksa	ime	prezima	adresa	šifra fakul.	broj računa	šifra tečaja	ocjena

*Slika 1.: Primjer loše strukturirane baze podataka*

Gornji primjer pokazuje anomalije sva tri tipa. Ako želimo dodati slog u bazi, a on se odnosi na studenta koji je uplatio neku od rata tečaja, onda moramo unositi podatke o fakultetu na kojem je student upisan. Dodati podatke o tečaju nije moguće sve dok student ne plati ratu za taj tečaj. Ubaciti podatke o fakultetu nije moguće sve dok neki student ne pohađa tečaj, a upisan je na taj fakultet (B.Markić (1996) "Kompjutori i informacijski sustavi").

Anomalija brisanja slogova. Ako brišemo podatke o studentu koji je jedini upisan na neki fakultet ili jedini polaže neki tečaj onda se gube podaci o fakultetu odnosno tečaju.

Anomalija izmjene sadržaja. Ponekad se naziva i potencijalna inkonzistentcija. Ako se promijeni naziv nekog tečaja onda se promjena mora izvršiti za sve kopije tog tečaja. Drugim riječima, promjena se mora izvršiti kod svih studenata koji su polaznici tog tečaja čije se ime mijenja. Realna je mogućnost da neka od kopija tečaja ostane neažurirana.

Sve ove anomalije su posljedica redudancije podataka. Model baze koji ne stvara anomalije ažuriranja (anomalije izmjene sadržaja, anomalije brisanja i anomalije dodavanja) naziva se kanonski model.

Osobine relacije zavise od semantičkih odnosa koji vrijede između atributa relacijske sheme. Postupak normalizacije, kojim se izbjegava višestruko zapisivanje istih podataka u bazi, prvi je predložio E. F. Codd za relacijski model podataka. Normalizacija je postupak u kome se na relacijskoj shemi koja nije normalizirana, uvode određena ograničenja. Transformacija relacijske sheme uzrokuje i transformaciju relacije.

Razlikujemo dvije metode normalizacije: **horizontalnu i vertikalnu**. Horizontalna normalizacija koristi operacije selekcije (operacija kojom se izdvajaju n-torke relacije) i unije. Atributi relacijske sheme se ne mijenjaju. Očekuje se njena primjena kod distribuiranih baza kod kojih cijela relacija ne mora biti memorirana na jednoj lokaciji (sadržaj jedne baze može biti memoriran na više lokacija) (detaljnije S.Tkalac (1994) "Relacijski model baza podataka").

Vertikalna normalizacija se sastoji u razdvajanju relacijske sheme (skupa atributa) na više manjih relacijskih shema (relacijske shema sa

manje atributa). Normalizacijom se ne smiju gubiti polazne relacijske sheme (relacijske sheme koja se normalizira). Mora se zadržati semantika polazne relacije. Općenito možemo reći da je izvršeno "dekomponiranje bez gubitka informacija" ako operacija prirodnog spajanja novih relacija po zajedničkom atributu daje polaznu relaciju (lossless join). Definiranje normalnih formi zahtjeva poznavanje pojmova funkcijske i višeznačne zavisnosti, te spajanja.

Prevođenjem relacija u više normalne forme mogu se prikazivati informacije koje se u nižim normalnim formama ne mogu vidjeti. Postoje i više normalne forme od treće normalne forme. Prikazat ćemo prvo tzv. BCNF normalnu formu forme.

### 2.3. Boyce Coddova normalna forma (BCNF)

Relacija se nalazi u Boyce-Coddovoj normalnoj formi (BCNF) ako je svaki determinant kandidat za ključ. Determinant je atribut o kojem je neki atribut potpuno funkcijski ovisan.<sup>1</sup> Codd-ova treća normalna forma je pokazala određene nedostatke. Nedostaci su vidljivi u relaciji koja ima više kandidata za ključ, a oni su složeni i imaju najmanje jedan zajednički atribut. Boyce i Codd su predložili novu normalnu formu, koju su nazvali Boyce-Coddova normalna forma. U našoj relaciji:

**student\_te(broj\_ind\_ime,prezime,adresafakul,adrf,šifra\_tečaja,ocjena)**

atributi **broj\_ind**, **fakul**, **šifra\_tečaja** i **(broj\_ind,šifra\_tečaja)** su determinanti.

BCNF se ne povezuje s prvom i drugom normalnom formom. Svaka relacija koja se nalazi u BCNF, se nalazi u trećoj normalnoj formi, ali ne vrijedi obratno.

Analizirajmo sljedeće primjere:

a) relacija

**student\_te (broj\_ind,ime,prezime,adresa,fakul,adrf,šifra\_tečaja,ocjena)**

nije u BCNF jer je od svih determinanti jedino atribut **(broj\_ind,šifra\_tečaja)** kandidat za ključ.

b) relacija:

**student (broj\_ind,ime,prezime,adresa,fakul,adrf)**

1 Postoji i druga definicija BCNF: "Neka je R relacijska shema i neka je skup atributa X podskup od R i neka atribut A je element razlike (R - X). Relacijska shema R se nalazi u BCNF ako se nalazi u prvoj normalnoj formi i ako za svako X koje funkcijski određuje A tj. X----- A vrijedi X----- R. Primjetimo da se gornja definicija ne razlikuje od prethodne date definicije. Skup atributa X je determinant jer atributi A funkcijski ovise o atributu X. Pošto skup atributa X funkcijski određuje R tj. X----- R znači da je determinant X kandidat za ključ relacije.

nije u BCNF jer su njene determinante **broj\_ind** i **fakul** a atribut **fakul** nije kandidat za ključ.

c) relacija:

**fak\_adresa (fakul,adrf)**

je u BCNF jer determinanta atribut **fakul** kandidat za ključ.

Prevođenje relacija u BCNF(35) se sastoji u određivanju determinanti - atributa od kojih drugi atributi potpuno funkcijski ovise i ispitivanju da li determinantni predstavljaju kandidati za ključeve.

### 2.3 Višeznačne funkcijske ovisnosti

BCNF je najviša normalna forma koja se može izvesti pomoću funkcijskih ovisnosti. Funkcijske ovisnosti nisu dovoljne da opišu semantiku veza entitet-atribut ili veza-atribut u relacijskom modelu. Primjer su situacije u kojima entitet može imati istodobno više vrijednosti nekog atributa. Zato je uveden pojam višeznačnih funkcijskih ovisnosti. Za razumijevanje prirode višeznačnih funkcijskih ovisnosti navest ćemo primjer.

projekt	izvođač	literatura
ekspertni	Brano	Lenat
sustavi	Marina	Feigenbaum
	Tanja	
jezici UI	Marko	Martin
		Morton

Za dati projekt postoji više izvođača i može se koristiti različita literatura u njegovoj izvedbi. Između izvođača na projektu i literature ne postoji funkcijska ovisnost. To znači da svaki izvođač projekta može koristiti bilo koju knjigu (literaturu). Za svaki projekt postoje sve moguće kombinacije bez ponavljanja, izvođača i literature. Općenito, ako u relaciji postoje n-torke:

$\langle P, II, LI \rangle$  i  $\langle P, I2, L2 \rangle$

moraju postojati n-torke

$\langle P, II, L2 \rangle$  i  $\langle P, I2, LI \rangle$

Na temelju opisane semantike modela, relaciju možemo prevesti u prvu normalnu formu (vi neključni atributi da funkcijski ovise o ključu)

samo ako navedemo sva kombinacije bez ponavljanja izvođača i dostupne im literature.

### PROJEKTI (PROJEKT,IZVOĐAČ,LITERATURA)

projekt	izvođač	literatura
eksp.sus.	Brano	Lenat
eksp.sus.	Brano	Feigenbaum
eksp.sus.	Marina	Lenat
eksp.sus.	Marina	Feigenbaum
eksp.sus.	Vanja	Lenat
eksp.sus.	Vanja	Feigenbaum
jeziciUI	Marko	Martin
jezici UI	Marko	Morton

Determinanta je atribut od kojeg neki funkcijski ovisi. U relaciji projekti - na temelju opisane semantike modela - postoji samo jedan determinat. To je skup (kolekcija) sva tri atributa. Uvjet bcnf, da je svaki determinat kandidat za ključ, je ispunjen te se relacija nalazi u BCNF:

Relacija pokazuje anomaliju ažuriranja. Na primjer, da bi smo dodali podatak: knjiga od Harmona se koristi za projekt ekspertni sustav (eksp.sus.), potrebno je dodati tri nove n-torke. Zato ćemo relaciju projekti dekomponirati u dvije relacije:

#### R1 (projekt, izvođač)

projekt	izvođač
ES	Brano
ES	Marina
Es	Tanja
UI	Marko

#### R2 (projekt, literatura)

projekt	literatura
ES	Lenat
ES	Feigenbaum
UI	Martin
UI	Morton

Dekomponiranje nismo izvršili na temelju normalnih formi, pomoću funkcijskih ovisnosti. U relacijama ne postoje funkcijske ovisnosti ali se mogu identificirati sljedeće semantičke ovisnosti:

- jedan projekt izvodi konačan skup izvođača,
- za jedan projekt se koristi određena literatura.

Ovakve ovisnosti se nazivaju višeznačne funkcijske ovisnosti. Na temelju datog primjera možemo dati formalniju definiciju višeznačnih funkcijskih ovisnosti:

U relaciji  $R(X, Y, Z)$  postoji višeznačna funkcijska ovisnost:

$X \dashrightarrow \dashrightarrow Y$  ako, i samo ako, kad god postoje u tijelu relacije  $n$ -torke

$(X, Y, z)$  i  $(x, y_1, z_1)$  postoje u tijelu relacije  $n$ -torke  $(x, y_1, z)$  i  $(x, y, z_1)$ .

Funkcijske ovisnosti koje smo ranije razmatrali su specijalni slučaj višeznačnih funkcijskih ovisnosti tj. onih funkcijskih ovisnosti za koje skup vrijednosti  $Y$ , za datu vrijednost atributa  $X$ , sadrži samo jednu vrijednost.

## ZAKLJUČAK

Danim primjerima normalizacije smo pokazali kako se eliminira višestruko zapisivanje istih podataka. Slaba strana normalizacije je, u pravilu, povećanje broja ulaznih i izlaznih operacija, što umanjuje brzine izračunavanja odgovorne upite. Normalizacija podataka barem do treće normalne forme je potrebna jer se smanjuje redundancija podataka, eliminira anomalija održavanja a baza podataka postaje "zajednička podatkovna osnova" budućih aplikacija.

## LITERATURA

Brano Markić (1996) "Kompjutori i informacijski sustavi"

Ekonomski fakultet Mostar.

Slavko Tkalec (1994) "Relacijski model baza podataka" DRIP, Zagreb.

David Vaskevite (1995) "Klijent/server strategije", Znak, Zagreb.

M.Varga., (1994) "Baza podataka: Konceptualno, logičko i fizičko modeliranje podataka, DRIP, Zagreb.

**Brano Markić, PhD**

Assistant Professor, Faculty of Economics, Mostar

**Ivan Pavlović, PhD**

Assistant Professor, Faculty of Economics, Mostar

## DATA NORMALIZATION AS A BASIC INDICATOR ON THE VALIDITY OF RELATIVE DATA MODELS

### Summary

*Decision-makers at various levels of organization, managers and beneficiaries, demand information in the carrying out of effective decisions. Reliable, relevant and timely information is a required condition for decision-making. We shall concentrate our attention on the organization of data, from which information is obtained. Generally, the concept of handling databases is the same for large and microcomputers.*

*Understanding the database system enables its' efficient use at the workstation. Previously, this kind of condition was not imposed upon the user. The development of powerful microcomputers and their networks imposed the task of understanding the organization and structure of databases. A prerequisite in the competent use of information in the computer age is the understanding of the principles of data files, databases, systems managing databases, data models and other principles needed for creating and discovering information.*

**Key words:** data models, normalization, normal forms