

OBLIKOVANJE RELACIJSKOG MODELA

III predavanje

Dr.sc. Emir Mešković

Kako oblikovati relacijski model?

- ▶ Izbor između nekoliko alternativnih modela
- ▶ CILJ – oblikovati model (shemu) baze podataka sa dobrim osobinama
- ▶ Karakteristike loše koncipiranog modela:
 - ▶ redundancija (čije su posljedice)
 - ▶ Neracionalno korištenje prostora za pohranu
 - ▶ Anomalija unosa
 - ▶ Anomalija izmjena
 - ▶ Anomalija brisanja
 - ▶ pojave lažnih n-torki



Primjer loše koncipirane sheme baze podataka

- ▶ Prodavnice šalju svoje narudžbe zastupniku:

<p>Bingo-Sjenjak I. Mujezinovića 24 75000 Tuzla</p> <p>Mepas d.o.o. Varaždinska 1 88220 Široki Brijeg</p> <p>Narudžba br 116/18 Datum: 10.10.2018.</p> <p>Molim isporučite nam 1500 komada Proizvoda Dado čokolada (šifra 111) 1150 komada proizvoda Medeno srce (šifra 222)</p>	<p>Interex-Luke I. Sarajlića 5 75000 Tuzla</p> <p>Mepas d.o.o. Varaždinska 1 88220 Široki Brijeg</p> <p>Narudžba br 8 Datum: 12.10.2018.</p> <p>Molim isporučite nam 1350 komada Proizvoda Dado čokolada 1400 komada proizvoda Munchmallow (šifra 333)</p>	<p>Bingo-Sjenjak I. Mujezinovića 24 75000 Tuzla</p> <p>Mepas d.o.o. Varaždinska 1 88220 Široki Brijeg</p> <p>Narudžba br 137/18 Datum: 15.10.2018.</p> <p>Molim isporučite nam 1600 komada Proizvoda Dado čokolada (šifra 111)</p>
---	---	---

- ▶ Zastupnik želi pohraniti podatke o narudžbama u svoju bazu podataka. Svi podaci se pohranjuju u relaciju prodaja

prodaja

- ▶ (nazPreduz, mjesto, adresa, brNarudz, sifArtikl, nazArtikl, kol, datIspor)

Neracionalno korišćenje prostora za pohranu

- ▶ Sadržaj relacije nakon unosa podataka iz prispjelih narudžbi:

Relacija *prodaja*

<i>(nazPreduz,</i>	<i>mjesto,</i>	<i>adresa,</i>	<i>brNarudz,</i>	<i>sifArtikl,</i>	<i>nazArtikl,</i>	<i>kol,</i>	<i>datIspor)</i>
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	111	Čokolada Dado	1500	10.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	222	Medeno srce	1150	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	111	Čokolada Dado	1350	12.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	333	Munchmallow	1400	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	137/18	111	Čokolada Dado	1600	15.10.

- ▶ Na više mjesta se ponavlja isti (redundantan) podatak
 - ▶ Bingo-Sjenjak je market u Tuzli
 - ▶ Adresa marketa Bingo-Sjenjak je I. Mujezinovića 24
 - ▶ Naziv artikla sa šifrom 111 je Čokolada Dado
 - ▶ Datum isporuke narudžbe 116/18 je 10.10.2018. godine



Anomalija unosa

Relacija *prodaja*

<i>(nazPreduz,</i>	<i>mjesto,</i>	<i>adresa,</i>	<i>brNarudz,</i>	<i>sifArtikl,</i>	<i>nazArtikl,</i>	<i>kol,</i>	<i>datIspor)</i>
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	111	Čokolada Dado	1500	10.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	222	Medeno srce	1150	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	111	Čokolada Dado	1350	12.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	333	Munchmallow	1400	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	137/18	111	Čokolada Dado	1600	15.10.

- ▶ Ne mogu se unijeti podaci o artiklima koji se trenutno ne prodaju
- ▶ Ne mogu se unijeti podaci o marketima kojima se trenutno ne prodaje
- ▶ ...
- ▶ Svaki put kada se unosi novi podatak o narudžbi nekog artikla, mora se ponovo upisati i naziv i mjesto i adresa marketa koji taj proizvod naručuje
 - ▶ Pri tome se mora paziti da se podaci za istu prodavnicu uvijek jednako unesu da bi se očuvala konzistentnost podataka



Anomalija izmjena

Relacija *prodaja*

<i>(nazPreduz,</i>	<i>mjesto,</i>	<i>adresa,</i>	<i>brNarudz,</i>	<i>sifArtikl,</i>	<i>nazArtikl,</i>	<i>kol,</i>	<i>datIspor)</i>
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	111	Čokolada Dado	1500	10.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	222	Medeno srce	1150	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	111	Čokolada Dado	1350	12.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	333	Munchmallow	1400	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	137/18	111	Čokolada Dado	1600	15.10.

- ▶ Ako neki market promijeni adresu, promjenu adrese potrebno je izvršiti u svim zapisima da bi se zadržala konzistentnost podataka
 - ▶ Npr. marketu Bingo-Sjenjak se adresa promijeni u Sjenjak B10/6

<i>(nazPreduz,</i>	<i>mjesto,</i>	<i>adresa,</i>	<i>brNarudz,</i>	<i>sifArtikl,</i>	<i>nazArtikl,</i>	<i>kol,</i>	<i>datIspor)</i>
Bingo-Sjenjak	Tuzla	Sjenjak B10/6	116/18	111	Čokolada Dado	1500	10.10.
Bingo-Sjenjak	Tuzla	Sjenjak B10/6	116/18	222	Medeno srce	1150	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	111	Čokolada Dado	1350	12.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	333	Munchmallow	1400	12.10.
Bingo-Sjenjak	Tuzla	Sjenjak B10/6	137/18	111	Čokolada Dado	1600	15.10.



Anomalija brisanja

Relacija *prodaja*

<i>(nazPreduz,</i>	<i>mjesto,</i>	<i>adresa,</i>	<i>brNarudz,</i>	<i>sifArtikl,</i>	<i>nazArtikl,</i>	<i>kol,</i>	<i>datIspor)</i>
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	111	Čokolada Dado	1500	10.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	222	Medeno srce	1150	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	111	Čokolada Dado	1350	12.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	333	Munchmallow	1400	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	137/18	111	Čokolada Dado	1600	15.10.

- ▶ **Brisanjem svih narudžbi o prodaji nekog artikla gube se podaci o artiklu**
 - ▶ Npr. Ako se obriše posljednja n-torka o narudžbama artikla Medeno srce podatke o tom artiklu više nećemo imati u bazi podataka

<i>(nazPreduz,</i>	<i>mjesto,</i>	<i>adresa,</i>	<i>brNarudz,</i>	<i>sifArtikl,</i>	<i>nazArtikl,</i>	<i>kol,</i>	<i>datIspor)</i>
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	116/18	111	Čokolada Dado	1500	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	111	Čokolada Dado	1350	12.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/18	333	Munchmallow	1400	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića 24	137/18	111	Čokolada Dado	1600	15.10.



Pokušaj (neuspješni) popravka sheme baze podataka

► Podaci o narudžbama će se pohranjivati u dvije relacije

► $narudzba = \pi_{nazPreduz, mjesto, adresa, brNarudz, datlspor, sifArtikl}(prodaja)$

► $artikl = \pi_{sifArtikl, nazArtikl, kol}(prodaja)$

narudzba

<i>(nazPreduz,</i>	<i>mjesto,</i>	<i>adresa,</i>	<i>brNarudz,</i>	<i>sifArtikl,</i>	<i>datlspor)</i>
Bingo-Sjenjak	Tuzla	I. Mujezinovića	116/12	111	10.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića	116/12	222	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/12	111	12.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/12	333	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića	137/12	111	15.10.

artikl

<i>(sifArtikl,</i>	<i>nazArtikl,</i>	<i>kol)</i>
111	Čokolada Dado	1500
222	Medeno srce	1150
111	Čokolada Dado	1350
333	Munchmallow	1400
111	Čokolada Dado	1600

Pojava lažnih n-torki

- ▶ Obavljanjem operacije *narudzba* $\triangleright \triangleleft$ *artikl* dobije se više n-torki nego ih je bilo u relaciji prodaja (neke n-torke u rezultatu su “lažne”)
 - ▶ *prodaja2* = *narudzba* $\triangleright \triangleleft$ *artikl* \neq *prodaja*

prodaja2

(nazPreduz,	mjesto,	adresa,	brNarudz,	sifArtikl,	nazArtikl,	kol,	datIspor)
Bingo-Sjenjak	Tuzla	I. Mujezinovića	116/12	111	Čokolada Dado	1500	10.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića	116/12	111	Čokolada Dado	1350	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića	116/12	111	Čokolada Dado	1600	15.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića	116/12	222	Medeno srce	1150	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/12	111	Čokolada Dado	1500	10.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/12	111	Čokolada Dado	1350	12.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/12	111	Čokolada Dado	1600	15.10.
Interex-Luke	Tuzla	I. Sarajlića 5	89/12	333	Munchmallow	1400	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića	137/12	111	Čokolada Dado	1500	10.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića	137/12	111	Čokolada Dado	1350	12.10.
Bingo-Sjenjak	Tuzla	I. Mujezinovića	137/12	111	Čokolada Dado	1600	15.10.

Ispravna shema baze podataka

market

(nazPreduz, mjesto, adresa,)

Bingo-Sjenjak	Tuzla	I. Mujezinovića 24
Interex-Luke	Tuzla	I. Sarajlića 5

artikl

(sifArtikl, nazArtikl,)

111	Čokolada Dado
222	Medeno srce
333	Munchmallow

narudzba

(nazPreduz, brNarudz, datIspor)

Bingo-Sjenjak	116/18	10.10.
Interex-Luke	89/18	12.10.
Bingo-Sjenjak	137/18	15.10.

stavkaNarudzbe

(brNarudz, sifArtikl, kol)

116/18	111	1500
116/18	222	1150
89/18	111	1350
89/18	333	1400
137/18	111	1600

► Za vježbu provjerite

- Postoji li redundancija u ovoj bazi podataka?
- Da li je moguća pojava lažnih n-torki?

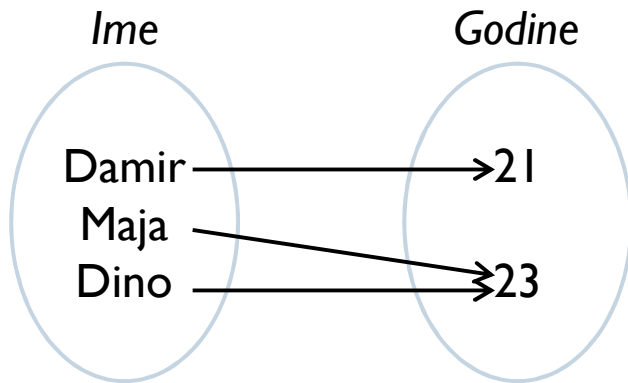


Kako odrediti zamjenu za loše koncipiranu relacijsku shemu?

- ▶ Proučavanjem značenja podataka (semantike)
- ▶ Proučavanjem zavisnosti među podacima
- ▶ Uvođenjem ograničenja koja su ovisna o semantici podataka
- ▶ Najvažnije su FUNKCIJSKE ZAVISNOSTI

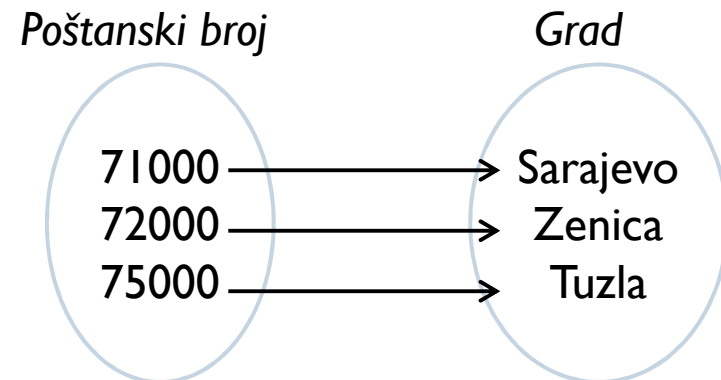
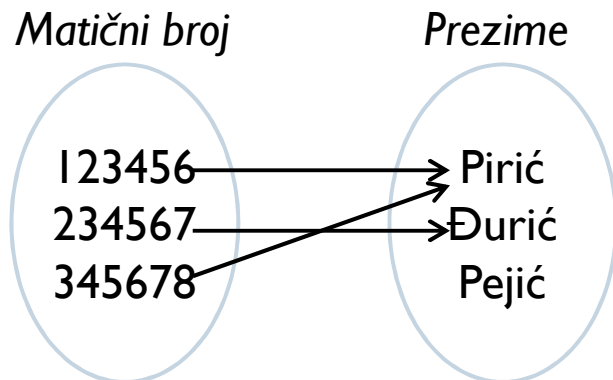


Funkcija



Preslikavanje kod kojeg vrijedi da je:

Svakom članu skupa *Ime* pridružen
jedan i samo jedan član skupa *Godine*



Funkcijske zavisnosti

- ▶ Neka su X i Y skupovi atributa
- ▶ Među njima postoji funkcijska zavisnost
- ▶ $X \rightarrow Y$
 - ▶ X funkcijski određuje Y
 - ▶ Y funkcijski ovisi o X
- ▶ ako je jednoj vrijednosti x od X u svakom trenutku pridružena jedna i samo jedna vrijednost y od Y



Funkcijske zavisnosti - definicija 1

- ▶ Neka je r relacija sa shemom R i neka su X i Y skupovi atributa, $X \subseteq R, Y \subseteq R$
- ▶ Relacija r zadovoljava funkcijsku zavisnost $X \rightarrow Y$ (funkcijska zavisnost $X \rightarrow Y$ vrijedi na shemi R):
- ▶ ako za svaku vrijednost x od X $\pi_Y(\sigma_{X=x}(r))$ ima najviše jednu n-torku



Funkcijske zavisnosti - primjer

osoba

<u>(matBr</u>	<u>prezime,</u>	<u>ime,</u>	<u>postBr,</u>	<u>nazivGrad)</u>
12345	Pirić	Damir	75000	Tuzla
23456	Đurić	Maja	71000	Sarajevo
34567	Pejić	Dino	72000	Zenica

Relacija *osoba* zadovoljava funkcijsku zavisnost $matBr \rightarrow prezime$ ako za svaku x vrijednost od *matBr* $\pi_{prezime}(\sigma_{matBr=x}(osoba))$ ima najviše jednu n-torku

$$\sigma_{matBr=12345}(osoba) = osoba_1$$

<u>(matBr</u>	<u>prezime,</u>	<u>ime,</u>	<u>postBr,</u>	<u>nazivGrad)</u>
12345	Pirić	Damir	75000	Tuzla

$$\pi_{prezime}(osoba_1) = osoba_2 \quad \begin{array}{c} \underline{(prezime)} \\ Pirić \end{array}$$



Funkcijske zavisnosti - primjer

osoba

<u>(matBr</u>	<u>prezime,</u>	<u>ime,</u>	<u>postBr,</u>	<u>nazivGrad)</u>
12345	Pirić	Damir	75000	Tuzla
23456	Đurić	Maja	71000	Sarajevo
34567	Pejić	Dino	72000	Zenica
45678	Pirić	Ema	75000	Tuzla

Relacija *osoba* zadovoljava funkcijsku zavisnost $postBr \rightarrow nazivGrad$ ako za svaku x vrijednost od $postBr$ $\pi_{nazivGrad}(\sigma_{postBr=x}(osoba))$ ima najviše jednu n-torku

$$\sigma_{postBr=75000}(osoba) = osoba_3$$

<u>(matBr</u>	<u>prezime,</u>	<u>ime,</u>	<u>postBr,</u>	<u>nazivGrad)</u>
12345	Pirić	Damir	75000	Tuzla
45678	Pirić	Ema	75000	Tuzla

$\pi_{grad}(osoba_3) = osoba_4$
 $\underbrace{(nazivGrad)}_{Tuzla}$

Funkcijske zavisnosti – definicija 2

- ▶ Neka je r relacija sa shemom R i neka su X i Y skupovi atributa, $X \subseteq R, Y \subseteq R$
- ▶ Relacija r zadovoljava funkcijsku zavisnost $X \rightarrow Y$ (funkcijska zavisnost $X \rightarrow Y$ vrijedi na shemi R):
- ▶ ako u svim dopuštenim stanjima relacije $r(R)$ svake dvije n-torke t_1 i t_2 sa jednakim X -vrijednostima imaju jednake Y -vrijednosti
- ▶ $t_1(X) = t_2(X) \Rightarrow t_1(Y) = t_2(Y)$



Funkcijske zavisnosti

- ▶ Funkcijske zavisnosti **proizilaze iz značenja podataka** (semantike), a ne iz trenutnog stanja relacije

<i>osoba</i>				
(matBr	<i>prezime,</i>	<i>ime,</i>	<i>postBr,</i>	<i>nazivGrad)</i>
12345	Pirić	Damir	75000	Tuzla
23456	Đurić	Maja	71000	Sarajevo
34567	Pejić	Dino	72000	Zenica

- ▶ Posmatranjem samo trenutnog stanja relacije mogli bismo (**pogrešno!**) zaključiti da vrijedi FZ *prezime* → *ime*
- ▶ Međutim, poznavanjem značenja podataka u relaciji *osoba* možemo zaključiti da je dopušteno unijeti n-torku <45678, Pirić, Ema, 75000, Tuzla>
- ▶ => FZ *prezime* → *ime* ne vrijedi u shemi *OSOBA* (relaciji *osoba*)



Priroda funkcijskih zavisnosti

- ▶ Postojanje funkcijske zavisnosti ne može se dokazati na temelju postojećih podataka u relaciji
- ▶ Analizom postojećih podataka u relaciji moguće je tek **pretpostaviti** da bi funkcijska zavisnost mogla vrijediti
- ▶ Dokaz za postojanje funkcijske zavisnosti treba tražiti u **značenju** pojedinih atributa



Priroda funkcijskih zavisnosti

r(R)		
A	B	C
a	α	1
b	β	1
b	α	1
c	α	3
a	α	1

r(R)		
A	B	C
a	α	1
b	β	1
b	α	1
c	α	3
a	α	1

- ▶ Vrijedi li FZ $AB \rightarrow C$ na shemi R ?
 - ▶ Moguće je da vrijedi, ali to ne možemo sa sigurnošću tvrditi
 - ▶ Bez poznavanja značenja atributa A , B i C , ne možemo zaključiti koje funkcijske zavisnosti zaista vrijede na shemi R
- ▶ Vrijedi li FZ $BC \rightarrow A$ na shemi R ?
 - ▶ Sa sigurnošću možemo tvrditi: NE

Priroda funkcijskih zavisnosti

- ▶ Ako u relacijskoj shemi R postoji funkcijska zavisnost $X \rightarrow Y$, tada relacija $r(R)$ ne može sadržavati dvije n -torke koje imaju jednake X -vrijednosti i različite Y -vrijednosti
- ▶ Primjer: ako u relacijskoj shemi
- ▶ $R = (\text{matBr}, \text{prezime}, \text{grad}, \text{telefon})$
- ▶ Postoji funkcijska zavisnost $\text{matBr} \rightarrow \text{prezime}$, tada relacija $r(R)$ ne može sadržavati dvije n -torke s istim matičnim brojem i različitim prezimenom

Priroda funkcijskih zavisnosti - primjer

- ▶ Studenta *mbrStud* je na ispitu iz predmeta *sifPred* na datum *datlspit* nastavnik *sifNast* ocijenio ocjenom *ocjena*

ispit				
mbrStud	sifPred	datlspit	sifNast	ocjena
1111	101	15.06.2017	31	5
1111	101	29.08.2017	32	8
2222	101	15.06.2017	30	7
2222	110	30.08.2017	32	6

- ▶ Vrijedi li FZ *mbrStud, sifNast* → *ocjena*
 - ▶ Ne, jer bi to značilo da nastavnik *x* studentu *y* uvijek mora dati istu ocjenu
- ▶ Vrijedi li FZ *mbrStud, sifPred* → *ocjena*
 - ▶ Ne, jer bi to značilo da student *x* iz predmeta *y* uvijek mora dobiti istu ocjenu
- ▶ Vrijedi li FZ *mbrStud, datlspit* → *ocjena*
 - ▶ Ne, jer bi to značilo da student *x* na datum *y* uvijek mora dobiti istu ocjenu
- ▶ Vrijedi li FZ *mbrStud, sifPred, sifNast* → *ocjena* NE (Zašto?)
- ▶ Vrijedi li FZ *mbrStud, sifPred, datlspit* → *ocjena* DA (Zašto?)

Funkcijske zavisnosti – SQL primjer

- ▶ Pomoću SELECT naredbe ispitati da li u relaciji ispit eventualno vrijedi funkcijska zavisnost $mbrStud, sifNast \rightarrow ocjena, datIspit$
- ▶ Ispituju se svi parovi n-torki t_1 i t_2 koje imaju jednake X -vrijednosti ($X = \{mbrStud, sifNast\}$)
- ▶ Ako postoji par n-torki t_1 i t_2 koje imaju iste X -vrijednosti , a različite Y -vrijednosti ($Y = \{ocjena, datIspit\}$) tada funkcijska zavisnost **sigurno ne vrijedi**

```
SELECT *  
  FROM ispit AS is1, ispit AS is2  
    WHERE is1.mbr = is2.mbr  
    AND is1.sifNast = is2.sifNast  
    AND (is1.ocjena <> is2.ocjena  
        OR is1.datIspit <> is2.datIspit)
```

n-torke t_1 i t_2
koje imaju jednake
 X -vrijednosti ...

... a različite
 Y -vrijednosti ...

- ▶ Ako takve n-torke ne postoje, onda FZ **možda** vrijedi
-

