

UVOD U BAZE PODATAKA

Vježbe 06

NORMALIZACIJA

Ukratko, **normalizacija** je proces efikasnog organiziranja podataka. To je postupak kojim se nenormalizirana relacijska shema transformira uvođenjem novih ograničenja.

Normalizacija za cilj ima dvije stvari:

- eliminaciju redundantnosti (npr. izbjeći spremanje istog podatka u dvije ili više tablica)
- osiguranje smisla međuovisnosti podataka (npr. spremanje samo povezanih podataka u tablicu).

Oba ova cilja smanjuju količinu prostora koju baza koristi i osiguravaju logičnost podataka.

NORMALIZACIJA

Osnovni razlog zašto se koristi normalizacija je taj što je ona vrlo moćno (i kroz praksu dokazano) *oružje* za borbu protiv moguće korupcije baze podataka.

Korupcija podataka nastaje u nekoliko slučajeva od kojih su tri posebno problematična i imaju svoja 'imena':

- Anomalije ubacivanja (*insertion anomalies*)
- Anomalije promjene podataka (*update anomalies*)
- Anomalije brisanja (*deletion anomalies*).

Promotrimo svaku od tih anomalija zasebno.

Anomalije ubacivanja

Pogledajmo kako je vlasnik videoteke "LošFilm" vodio svoje poslovanje prije nego je otkrio (relacijske) baze podataka:

Članski broj	Ime i prezime	Adresa	Telefon	Film	Datum posudbe	Datum vraćanja	Naplaćena zakasnina
21	Katarina Gjalski	Kneza Mislava 11a	085-654321	Matrix 1	15.01.09.	17.01.09.	5,00 kn
22	Nikola Jurčević	Zagorska 194	086-123145	Matrix 1	18.01.09.	19.01.09.	- kn
22	Nikola Jurčević	Zagorska 194	086-123245	Matrix 2	19.01.09.	22.01.09.	10,00 kn
21	Katarina Gjalski	Kneza Mislava 211b	085-654321	One Flew Over the Cuckoo's Nest	19.01.09.	20.01.09.	- kn
21	Katarina Gjalski	Kneza Višeslava 11a	085-654321	My Man Godfrey	20.01.09.	22.01.09.	5,00 kn
42	Stjepan Kotoromanić	Ilica 224	084-531514	Pulp Fiction	11.02.09.	12.02.09.	- kn
22	Nikola Jurčević	Zagorska 194	086-123155	One Flew Over the Cuckoo's Nest	19.01.09.	20.01.09.	- kn

Anomalije kod ubacivanja se pojavljuju onog trenutka kada se upisivanjem u tablicu podaci krivo napišu: gdje točno živi Katarina Gjalski? Koji je broj telefona Nikole Jurčevića?

Ti podaci su prilikom zapisivanja krivo uneseni i stoga su konzistentnost i integritet podataka narušeni, tj. **više ne postoji točna informacija.**

Anomalije brisanja

Anomalije brisanja nastaju kada brisanjem jednih podataka, obrišemo i neke druge koji nisu direktno povezani:

Članski broj	Ime i prezime	Adresa	Telefon	Film	Datum posudbe	Datum vraćanja	Naplaćena zakasnina
21	Katarina Gjalski	Kneza Mislava 11a	085-654321	Matrix 1	15.01.09.	17.01.09.	5,00 kn
22	Nikola Jurčević	Zagorska 194	086-123145	Matrix 1	18.01.09.	19.01.09.	- kn
22	Nikola Jurčević	Zagorska 194	086-123245	Matrix 2	19.01.09.	22.01.09.	10,00 kn
21	Katarina Gjalski	Kneza Mislava 211b	085-654321	One Flew Over the Cuckoo's Nest	19.01.09.	20.01.09.	- kn
21	Katarina Gjalski	Kneza Višeslava 11a	085-654321	My Man Godfrey	20.01.09.	22.01.09.	5,00 kn
42	Stjepan Kotoromanić	Ilica 224	084-531514	Pulp Fiction	11.02.09.	12.02.09.	- kn
22	Nikola Jurčević	Zagorska 194	086-123155	One Flew Over the Cuckoo's Nest	19.01.09.	20.01.09.	- kn

Brisanjem posudbe za film Pulp Fiction, ostajemo bez podataka o Stjepanu Kotoromaniću! Više nemamo njegovu adresu, telefon... zapravo, niti ne znamo da je on član naše videoteke.

Kao i u prethodnom slučaju, nakon ovakvog brisanja, **više ne postoji točna informacija.**

Anomalije promjene

Anomalije promjene podataka nastaju prilikom promjene nekog podatka u nenormaliziranoj tablici:

Članski broj	Ime i prezime	Adresa	Telefon	Film	Datum posudbe	Datum vraćanja	Naplaćena zakasnina
21	Katarina Gjalski	Kneza Mislava 11a	085-654321	Matrix 1	15.01.09.	17.01.09.	5,00 kn
22	Nikola Jurčević	Pantovčak 42	086-123145	Matrix 1	18.01.09.	19.01.09.	- kn
22	Nikola Jurčević	Pantovčak 42	086-123245	Matrix 2	19.01.09.	22.01.09.	10,00 kn
21	Katarina Gjalski	Kneza Mislava 211b	085-654321	One Flew Over the Cuckoo's Nest	19.01.09.	20.01.09.	- kn
21	Katarina Gjalski	Kneza Višeslava 11a	085-654321	My Man Godfrey	20.01.09.	22.01.09.	5,00 kn
42	Stjepan Kotoromanić	Ilica 224	084-531514	Pulp Fiction	11.02.09.	12.02.09.	- kn
22	Nikola Jurčević	Zagorska 194	086-123155	One Flew Over the Cuckoo's Nest	19.01.09.	20.01.09.	- kn

Npr. ako je Nikola Jurčević promijenio adresu i sada živi na Pantovčaku 42, a kako njegova adresa u tablici piše na tri mjesta, lako se dogodi da se promjena ne učini na sva tri mjesta, pa opet nemamo konzistentne podatke.

Nakon ovakvih izmjena, **više ne postoji točna informacija**, niti se može lako saznati koja od informacija je točna (živi li Nikola na Pantovčaku ili u Zagorskoj?).

NORMALIZACIJA

Svi ovi problemi mogu se riješiti dovođenjem relacija u odgovarajuće **normalne forme (NF)**:

- 1NF – vrijednost svakog atributa je jednostruka i nedjeljiva
- 2NF – svaki ne-primarni atribut ovisi samo o cijelom primarnom ključu
- 3NF – nema tranzitivnih ovisnosti

Postoje još nekoliko normalnih formi: *Boyce-Codd*-ova normalna forma (BCNF), 4NF, 5NF, DKNF (*Domain/key normal form*) i 6NF. Na vježbama ćemo govoriti samo o prvim trima formama.

Neka “pravila” koja nam mogu pomoći prilikom transformacije relacije su sljedeća:

- 1NF - izbaci ponavljajuće grupe podataka
- 2NF - izbaci sve što ne ovisi o cijelom primarnom ključu
- 3NF - izbaci stupce koji ovise o primarnom ključu i o još nečemu

1. normalna forma

Članski broj	Ime i prezime	Adresa	Telefon	Film1	Film2	Film3
21	Katarina Gjalski	Kneza Mislava 11a	085-654321	Matrix 1	One Flew Over the Cuckoo's Nest	My Man Godfrey
22	Nikola Jurčević	Zagorska 194	086-123145	Matrix 1	Matrix 2	One Flew Over the Cuckoo's Nest
42	Stjepan Kotoromanić	Ilica 224	084-531514	Pulp Fiction	Serenity	

Članski broj	Ime i prezime	Adresa	Telefon1	Telefon2
21	Katarina Gjalski	Kneza Mislava 11a	085-654321	018-111111
22	Nikola Jurčević	Zagorska 194	086-123145	061-123145
42	Stjepan Kotoromanić	Ilica 224	084-531514, 084-123124, 045-456465	

1NF zahtijeva da se grupe podataka ne ponavljaju. Npr. stupci Film1, Film2, Film3, ... Što ako netko posudi 50 filmova!? Tablica bi trebala sadržavati 50 Film stupaca! Očito je to nešto što ne želimo. Slično vrijedi i za telefon, a posebno nije dobro imati u jednom zapisu više informacija, kao što je to telefon Stjepana Kotoromanića. Postavljanje upita bi bilo vrlo nezgodno.

1. normalna forma

Prva normalna forma zahtijeva da te podatke prebacimo u više tablica.

Izbaciti ponavljajuće grupe podataka

Popis posudbi članova						
Članski broj	Ime i prezime	Adresa	Telefon	Film1	Film2	Film3
21	Katarina Gjalski	Kneza Mislava 11a	085-654321	Matrix 1	One Flew Over the Cuckoo's Nest	My Man Godfrey
22	Nikola Jurčević	Zagorska 194	086-123145	Matrix 1	Matrix 2	One Flew Over the Cuckoo's Nest
42	Stjepan Kotoromanić	Ilica 224	084-531514	Pulp Fiction	Serenity	

Popis članova			
Članski broj	Ime i prezime	Adresa	Telefon
21	Katarina Gjalski	Kneza Mislava 11a	085-654321
22	Nikola Jurčević	Zagorska 194	086-123145
42	Stjepan Kotoromanić	Ilica 224	084-531514



Popis posuđenih filmova			
IDPosudba	Članski broj	Film	Trajanje filma
1	21	Matrix 1	139
2	21	One Flew Over the Cuckoo's Nest	120
3	21	My Man Godfrey	125
4	22	Matrix 1	139
5	22	Matrix 2	145
6	22	One Flew Over the Cuckoo's Nest	120
7	42	Pulp Fiction	145
8	42	Serenity	155



2. normalna forma

U 2NF svi stupci koji nisu ključevi moraju ovisiti samo o ***cijelom*** ključu.

Normalne forme su progresivne: ***kako bi mogli postići drugu normalnu formu, relacija već mora biti u 1NF.***

Izbaciti sve što ne ovisi o cijelom primarnom ključu

2. normalna forma

Pogledajmo popis posuđenih filmova. Uočavamo probleme:

- Ako član 42 vrati film "Pulp Fiction" (to zabilježimo tako da obrišemo taj redak), podaci o tom filmu nestaju.
- Promjene podataka o filmu moramo raditi na više mjesta.

Niti jedan stupac zasebno nije kandidat za ključ (jer želimo da isti član posudi više filmova i da neki film može posuditi više članova).

Zajedno, članski broj i film daju kandidata za primarni ključ. Stupac koji preostaje (trajanje filma) tada ovisi samo o dijelu ključa (naziv filma).

Popis posuđenih filmova		
Članski broj	Film	Trajanje filma
21	Matrix 1	139
21	One Flew Over the Cuckoo's Nest	120
21	My Man Godfrey	125
22	Matrix 1	139
22	Matrix 2	145
22	One Flew Over the Cuckoo's Nest	120
42	Pulp Fiction	145
42	Serenity	155

To znači da je ova tablica podložna anomalijama prilikom promjene podataka.

2. normalna forma

Izbaciti sve što ne ovisi o cijelom primarnom ključu

Popis posuđenih filmova		
Članski broj	Film	Trajanje filma
21	Matrix 1	139
21	One Flew Over the Cuckoo's Nest	120
21	My Man Godfrey	125
22	Matrix 1	139
22	Matrix 2	145
22	One Flew Over the Cuckoo's Nest	120
42	Pulp Fiction	145
42	Serenity	155

Film				
IDFilm	Film	Trajanje	Glumci	Opis
1	Matrix 1	139
2	One Flew Over the Cuckoo's Nest	120
3	My Man Godfrey	125
4	Matrix 2	145
5	Pulp Fiction	145
6	Serenity	155



Posudbe			
FilmID	ClanID	Datum Posudbe	Datum Vracanja
1	21	15.01.09.	17.01.09.
1	22	18.01.09.	19.01.09.
4	22	19.01.09.	22.01.09.
2	21	19.01.09.	20.01.09.
3	21	20.01.09.	22.01.09.
5	42	11.02.09.	12.02.09.

3. normalna forma

Normalne forme su progresivne: *kako bi mogli postići treću normalnu formu, tablica već mora biti u 2NF.*

Tablica je u 3NF ako je u 2NF i ako vrijedi: svaki stupac ovisi samo o primarnom ključu. Obično tablica nije u 3NF kada jedan stupac ovisi u drugom, a taj onda ovisi o primarnom ključu. To nazivamo **tranzitivna ovisnost**.

Izbaciti stupce koji ne ovise o primarnom ključu

Popis članova					
IDČlan	Ime i prezime	Adresa	Telefon	Poštanski broj	Mjesto
21	Katarina Gjalski	Kneza Mislava 11a	085-654321	10000	Zagreb
22	Nikola Jurčević	Zagorska 194	086-123145	44000	Sisak
42	Stjepan Kotoromanić	Ilica 224	084-531514	21000	Split

3. normalna forma

Popis članova					
IDClan	Ime i prezime	Adresa	Telefon	Poštanski broj	Mjesto
21	Katarina Gjalski	Kneza Mislava 11a	085-654321	10000	Zagreb
22	Nikola Jurčević	Zagorska 194	086-123145	44000	Sisak
42	Stjepan Kotoromanić	Illica 224	084-531514	21000	Split

U popisu članova, mjesto ovisi o poštanskom broju koji pak ovisi o članskom broju. Dovoljno je u ovoj tablici zapamtiti samo poštanski broj (ili još bolje, kao strani ključ u tablicu s mjestima i poštanskim brojevima). Npr:

Član(Članski broj, Ime i prezime, Adresa, Telefon, Poštanski broj)

Mjesto(Poštanski broj, Naziv)

Drugo moguće rješenje je napraviti još jednu tablicu koja veže člana i mjesto, npr:

Član(Članski broj, Ime i prezime, Adresa, Telefon)

Mjesto(Poštanski broj, Naziv)

MjestoČlana(Članski broj, Poštanski broj)

Zadatak 1: Normalizirajte tablicu

Ime	Prezime	Adresa	Poštanski broj	Mjesto	Poduzeće	Url1	Url2	Url3	Url4
Pero	Perić	Unska 16	10000	Zagreb	Algebra d.o.o.	www.algebra.hr	www.racunarstvo.hr	databases.about.com	www.datamodel.org
Ana	Milić	Zagrebačka 28	10000	Zagreb	Algebra d.o.o.	www.sqlite.org	www.firebirdsql.org	www.racunarstvo.hr	www.mysql.com
Sanja	Tarak	Anićeva 42	21000	Split	LošFilm	www.racunarstvo.hr	www.datamodel.org	www.sqlite.org	www.readwriteweb.com/archives/rel_database.php
Nikola	Jurčević	Zagorska 194	10000	Zagreb	Algebra d.o.o.	www.sqlite.org	www.zagreb.hr	www.algebra.hr	www.joelonsoftware.com
Katarina	Gjalski	Kneza Mislava 11a	43000	Sisak	LošFilm	www.mysql.com	www.joelonsoftware.com	databases.about.com	www.firebirdsql.org
Stjepan	Kotoromanić	Ilica 224	21000	Split	LošFilm	www.racunarstvo.hr	www.datamodel.org	www.joelonsoftware.com	www.sqlite.org
Ivan	Mažuranić	Mažuranićeva 11	43000	Sisak	Algebra d.o.o.	www.firebirdsql.org	www.algebra.hr	www.datamodel.org	www.mysql.com

Zadatak 2: Normalizirajte tablicu

Datum	Kupac	Dućan	Prodavač	Artikl 1	Količina	Artikl 1	Količina	Artikl 3	Količina	Artikl 4	Količina
15.01.09.	Katarina Gjalski	Kneza Mislava 11a	Pero	Bijeli kruh	1	Hren 250 g	2	Mineralna voda 1 l	1	Obična voda 0.5 l	2
18.01.09.	Nikola Jurčević	Zagorska 194	Petar	Mlijeko 1 l	3	Hren 250 g	1	Obična voda 0.5 l	3		
19.01.09.	Nikola Jurčević	Zagorska 194	Pero	10 jaja	2	Mlijeko 1 l	1				
19.01.09.	Katarina Gjalski	Kneza Mislava 211b	Krešo	Hren 250 g	1	Mlijeko 1 l	5	Pivo 0.5 l	1		

Zadatak 3: Normalizirajte tablicu

Datum	Vlasnik auta	Adresa servisa	Mehaničar	Dio 1	Količina	Dio 2	Količina	Dio 3	Količina	Dio 4	Količina	Ukupna cijena
15.01.09.	Katarina Gjalski	Kneza Mislava 11a	Pero	Filtar ulja	1	Ulje 1 l	5	Svjećice	4	Filtar zraka	1	2,500.00 kn
18.01.09.	Nikola Jurčević	Zagorska 194	Petar	Filtar klime	1	Ulje 1 l	4					1,050.00 kn
19.01.09.	Nikola Jurčević	Zagorska 19 4	Pero	Metlice brisača	3							200.00 kn
19.01.09.	Katarina Gjalski	Kneza Mislava 211b	Krešo	Prednji defni far	1	Prednji desni blatobran	1	Prednja desna poluosovina	1	Felga	1	7,800.00 kn