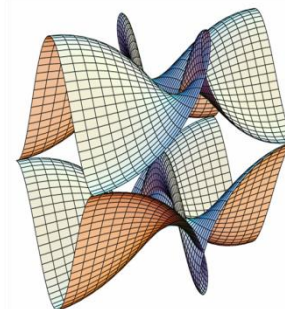




Sveučilište u Zagrebu
PMF – Matematički odsjek

BAZE PODATAKA
Predavanja 2019/2020



Poglavlje 4:

Normalizacija – nastavak

logičkog oblikovanja baze

podataka

Sastavio: Robert Manger
30.03.2020

Općenito o normalizaciji (1)

- Nastavljamo govoriti o drugoj fazi oblikovanja baze podataka - logičko oblikovanje.
- Opisat ćemo postupak daljnjeg dotjerivanja ili *normalizacije* polazne relacijske sheme.
- Dotjerivanje je potrebno zato što polazna relacijska shema može sadržavati nepravilnosti.
- Te nepravilnosti treba otkloniti prije nego što krenemo u fizičko oblikovanje.

Općenito o normalizaciji (2)

- Najprije opisujemo jednostavniji i u praksi više rabljeni dio normalizacije:
 - prevođenje u prvu, drugu i treću normalnu formu.
- Zatim se bavimo složenijim dijelom normalizacije koji je vezan uz:
 - Boyce-Codd-ovu i četvrtu normalnu formu.
- Na kraju raspravljamo:
 - zašto je normalizacija uopće potrebna, možemo li od nje odustati.

Sadržaj Poglavlja 4

4.1. Prva, druga i treća normalna forma

4.2. Boyce-Codd-ova i četvrta normalna forma

4.3. Potreba za normalizacijom

Općenito o prve tri normalne forme

- Teorija normalizacije zasnovana je na pojmu *normalnih formi*.
 - Svaka normalna forma predstavlja „zahtjev na kvalitetu“ relacije.
 - Što je normalna forma viša, zahtjevi su stroži.
 - Relacije dobivene postupkom iz Poglavlja 2 i 3 morale bi biti barem u prvoj normalnoj formi.
 - No Edgar Codd je u ranim 1970-tim godinama definirao drugu i treću normalnu formu.

Prevođenje u prvu normalnu formu (1)

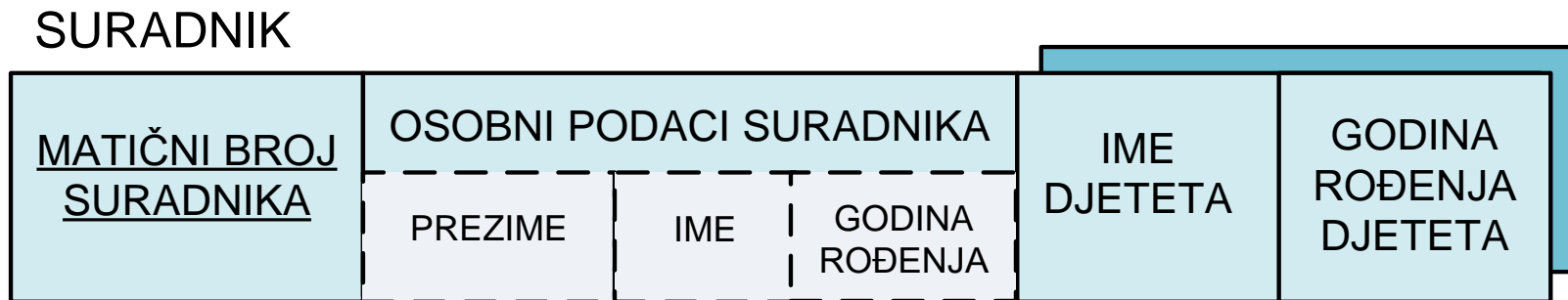
- Vrijednost atributa unutar relacije mora biti *jednostruka i jednostavna*. Ovo svojstvo zove se svojstvo *prve normalne forme* (oznaka: 1NF).
- 1NF zapravo ne predstavlja nikakav posebni zahtjev na relaciju,
 - To svojstvo već je ugrađeno u sam relacijski model.
 - U relacijskoj bazi podataka ne može postojati relacija koja ne bi bila u 1NF.
- Pojam 1NF izmišljen je zbog drugih modela podataka gdje podaci ne moraju biti normalizirani.

Prevođenje u prvu normalnu formu (2)

- Ako bazu oblikujemo u skladu s Poglavljima 2 i 3, tada se nećemo susresti se s nenormaliziranim podacima.
- Znanje o 1NF potrebno nam je kad neke podatke već imamo pohranjene u nekom drugom obliku te ih želimo prebaciti u relacijsku bazu.
- Zapisi (slogovi) u ne-relacijskim kolekcijama podataka mogu sadržavati:
 - *pod-zapise,*
 - *ponavljajuće skupine.*

Prevođenje u prvu normalnu formu (3)

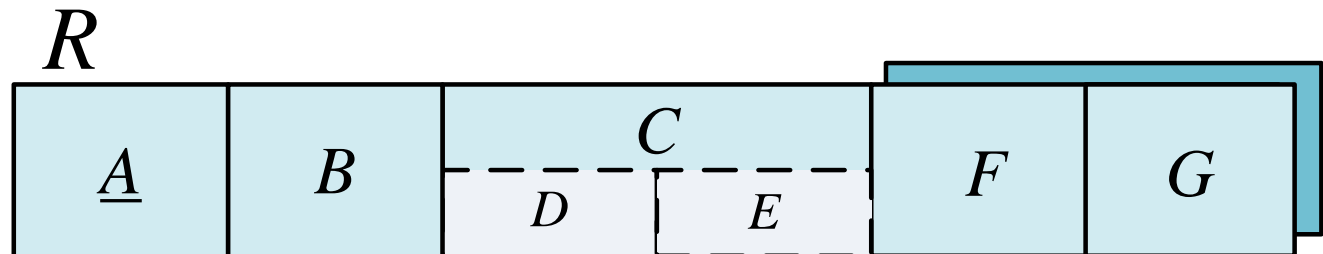
- U poduzeću imamo datoteku s podacima o suradnicima.
- Zapis o jednom suradniku sadrži:
 - Pod-zapis s osobnim podacima samog suradnika.
 - Ponavljajuću skupinu koja se ponavlja za svako dijete tog suradnika.
- Podaci očito nisu u 1NF.



Prevođenje u prvu normalnu formu (4)

- Da bismo nenormalizirane podatke pohranili u relacijskoj bazi, moramo ih prevesti u 1NF.
- Postupak prevođenja:
 - Uvodi se jedna osnovna relacija, te onoliko pomoćnih relacija koliko ima ponavljajućih skupina.
 - Fiksni dio zapisa prikazuje se kao jedna n -torka u osnovnoj relaciji.
 - Svaki pod-zapis rastavljen na nekoliko zasebnih atributa.
 - Svaka pojava ponavljajuće skupine prikazuje se kao zasebna n -torka u odgovarajućoj pomoćnoj relaciji.
 - Zbog čuvanja veze s polaznim zapisom u tu n -torku prepisuju se neki od identifikacijskih podataka iz fiksnog dijela zapisa.

Prevođenje u prvu normalnu formu (5)



- Shematski prikaz prevođenja u 1NF:

$R (\underline{A}, B, C-D, C-E)$
 $S (\underline{A}, \underline{E}, G)$

Prevođenje u prvu normalnu formu (6)

- Ako postupak prevođenja u 1NF primijenimo na naše zapise o suradnicima, dobit ćemo relacije:

SURADNIK (MATIČNI BROJ SURADNIKA, PREZIME SURADNIKA,
IME SURADNIKA, GODINA ROĐENJA SURADNIKA)

DIJETE (MATIČNI BROJ SURADNIKA, IME DJETETA,
GODINA ROĐENJA DJETETA) .

- Podaci o suradniku koji ima troje djece prikazani su jednom n -torkom u relaciji SURADNIK i trima n -torkama u relaciji DIJETE (po jedna za svako dijete).
- Veza između te četiri n -torke uspostavlja se na osnovi iste vrijednosti MATIČNOG BROJA SURADNIKA.

Funkcionalne ovisnosti (1)

- Većina normalnih formi zasnovana je na pojmu funkcionalne ovisnosti.
- Za relaciju R , atribut B od R je *funkcionalno ovisan* o atributu A od R (oznaka: $A \rightarrow B$) ako vrijednost od A jednoznačno određuje vrijednost od B .
 - Ako u R postoje dvije n -torke s jednakom vrijednošću od A , tada te n -torke moraju imati jednaku vrijednost od B .
 - Analogna definicija vrijedi i kad su A i B složeni atributi (skupine atributa).

Funkcionalne ovisnosti (2)

- Promatramo sljedeću (loše oblikovanu) relaciju:

UPISAO (JMBAG, ŠIFRA PREDMETA, NASLOV PREDMETA,
OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA, OCJENA) .

- JMBAG jednoznačno određuje studenta.
 - ŠIFRA PREDMETA jednoznačno određuje predmet.
 - Svaki predmet ima jednog nastavnika.
 - Svaki nastavnik ima jednu sobu.
 - OIB NASTAVNIKA jednoznačno određuje nastavnika.
- U relaciji postoji velik broj funkcionalnih ovisnosti:

Funkcionalne ovisnosti (3)

(JMBAG, ŠIFRA PREDMETA) → OCJENA ,
ŠIFRA PREDMETA → NASLOV PREDMETA ,
ŠIFRA PREDMETA → OIB NASTAVNIKA ,
ŠIFRA PREDMETA → BROJ SOBE NASTAVNIKA ,
OIB NASTAVNIKA → BROJ SOBE NASTAVNIKA .

- Za relaciju R , atribut B od R je *potpuno funkcionalno ovisan* o (složenom) atributu A od R ako vrijedi: B je funkcionalno ovisan o A , no B nije funkcionalno ovisan ni o jednom pravom podskupu od A .

Funkcionalne ovisnosti (4)

- Važna funkcionalna ovisnost koju redovito susrećemo u relaciji je *ovisnost atributa o ključu*.
- Svaki atribut relacije je funkcionalno ovisan o ključu, no ta ovisnost ne mora biti potpuna.
 - U prethodnoj relaciji UPISAO, atribut OCJENA je potpuno funkcionalno ovisan o primarnom ključu (JMBAG, ŠIFRA PREDMETA).
 - Atributi NASLOV PREDMETA, OIB NASTAVNIKA, te BROJ SOBE NASTAVNIKA nisu potpuno funkcionalno ovisni o primarnom ključu, budući da su ovisni samo o ŠIFRA PREDMETA, a ne i o JMBAG-u.

Prevođenje u drugu norm formu (1)

- U relaciji UPISAO postoje atributi koji nisu potpuno funkcionalno ovisni o primarnom ključu. Za njih kažemo da su *parcijalno ovisni* o tom ključu.
 - Parcijalna ovisnost smatra se nepoželjnim svojstvom. Zato se uvodi pojam druge normalne forme.
- Relacija je u *drugoj normalnoj formi* (oznaka: 2NF) ako je svaki njezin ne-primarni atribut potpuno funkcionalno ovisan o primarnom ključu.
 - Relacija je u 2NF ako u njoj nema parcijalnih ovisnosti atributa o primarnom ključu.

Prevođenje u drugu norm formu (2)

- Definicija 2NF ima smisla jedino ako je primarni ključ složen.
 - Relacija s jednostavnim ključem automatski je u 2NF.
- Relacija UPISAO nije u 2NF jer u njoj postoje parcijalne ovisnosti:
 - ŠIFRA PREDMETA → NASLOV PREDMETA ,
 - ŠIFRA PREDMETA → OIB NASTAVNIKA ,
 - ŠIFRA PREDMETA → BROJ SOBE NASTAVNIKA .

Prevođenje u drugu norm formu (3)

- Relacija koja nije u 2NF loše je oblikovana te se mora prevesti u 2NF.
- Postupak prevođenja:
 - Uz polaznu relaciju dodaje se onoliko novih relacija koliko ima različitih dijelova primarnog ključa koji sudjeluju u parcijalnim ovisnostima.
 - Iz polazne relacije *izbacuju se i prebacuju* u nove svi oni atributi koji su parcijalno ovisni o ključu.
 - Pritom u jednu novu relaciju idu oni atributi koji su ovisni o istom dijelu primarnog ključa.
 - Uz prebačene attribute, u novu relaciju *prepisuje se i odgovarajući dio primarnog ključa*, te on postaje ključ u toj novoj relaciji.

Prevođenje u drugu norm formu (4)

- U našem primjeru, prevođenje u 2NF razbit će polaznu relaciju UPISAO u dvije relacije:

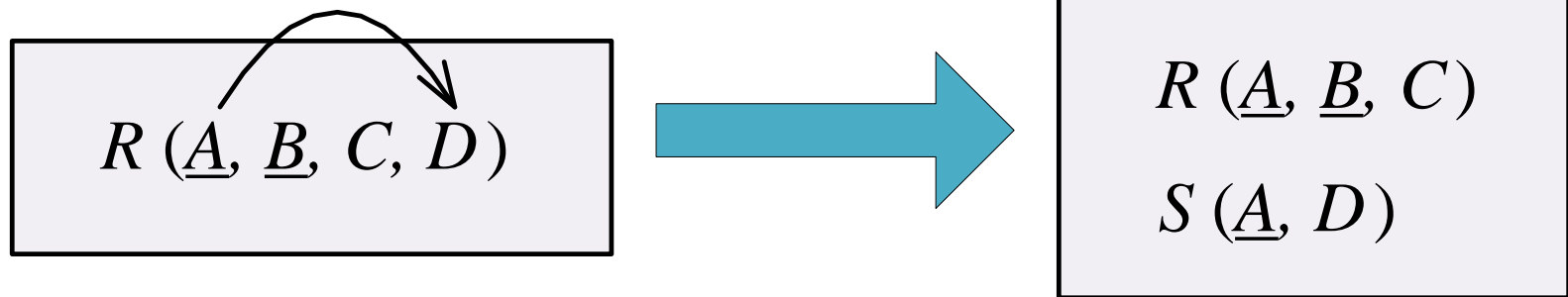
UPISAO (JMBAG, ŠIFRA PREDMETA, OCJENA)

PREDMET (ŠIFRA PREDMETA, NASLOV PREDMETA,
OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA) .

- U polaznoj verziji UPISAO sve parcijalne ovisnosti kreću iz ŠIFRE PREDMETA.
- Postupak prevođenja stvara samo jednu novu relaciju, u nju su prebačeni svi parcijalno-ovisni atributi zajedno sa ŠIFROM PREDMETA.
- ŠIFRA PREDMETA postaje ključ u novoj relaciji.

Prevođenje u drugu norm formu (5)

- Nakon ovog prevođenja, obje relacije su u 2NF.
 - U novoj verziji UPISAO više nema parcijalnih ovisnosti.
 - PREDMET ima jednostavan ključ pa je automatski u 2NF.
- Shematski prikaz prevođenja u 2NF:



Prevođenje u treću norm formu (1)

- U prethodnoj relaciji PREDMET postoji sljedeći niz funkcionalnih ovisnosti:
ŠIFRA PREDMETA → OIB NASTAVNIKA → BROJ SOBE NASTAVNIKA .
- Ovakav niz, pod uvjetom da srednji atribut nije kandidat za ključ, nazivamo *tranzitivna* ovisnost.
 - BROJ SOBE NASTAVNIKA je tranzitivno ovisan o ŠIFRI PREDMETA, posredstvom OIB NASTAVNIKA.
 - Pritom OIB NASTAVNIKA nije kandidat za ključ.
 - Tranzitivna ovisnost se smatra nepoželjnim svojstvom. Zato se uvodi pojam treće normalne forme.

Prevođenje u treću norm formu (2)

- Relacija je u *trećoj normalnoj formi* (oznaka: 3NF) ako je u 2NF i ako ne sadrži tranzitivne ovisnosti.
 - Relacija R je u 3NF ako za svaku funkcionalnu ovisnost $X \rightarrow A$ u R , takvu da A nije dio od X , vrijedi: X sadrži ključ za R ili je A primarni atribut.

- Relacija PREDMET nije u 3NF jer imamo funkcionalnu ovisnost

OIB NASTAVNIKA \rightarrow BROJ SOBE NASTAVNIKA .

- Pritom OIB NASTAVNIKA nije ključ, a BROJ SOBE NASTAVNIKA nije primarni atribut.

Prevođenje u treću norm formu (3)

- Relacija koja nije u 3NF loše je oblikovana, te se preporuča da se ona prevede u 3NF.
- Postupak prevođenja:
 - Uz polaznu relaciju dodaje se onoliko novih relacija koliko ima različitih atributa koji se pojavljuju kao srednji atributi u tranzitivnim ovisnostima.
 - Iz polazne relacije *izbacuju se i prebacuju* u nove svi oni atributi koji su tranzitivno ovisni.
 - Pritom u jednu novu relaciju idu oni atributi u čijim tranzitivnim ovisnostima se pojavljuje isti srednji atribut.
 - Uz prebačene attribute, u novu relaciju *prepisuje se i odgovarajući srednji atribut*, te on postaje ključ u toj novoj relaciji.

Prevođenje u treću norm formu (4)

- U našem primjeru, prevođenje u 3NF razbija relaciju PREDMET u sljedeće dvije relacije:

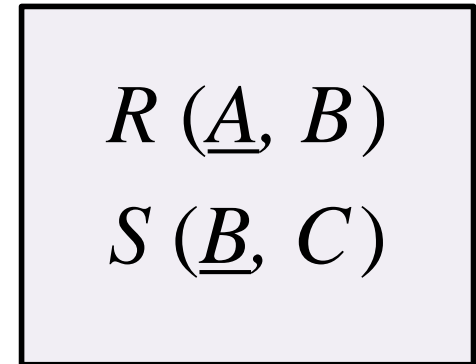
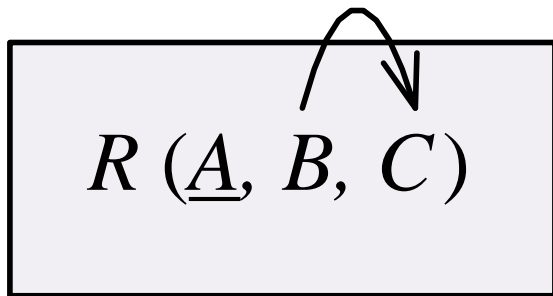
PREDMET (ŠIFRA PREDMETA, NASLOV PREDMETA,
OIB NASTAVNIKA)

NASTAVNIK (OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA) .

- U polaznoj verziji od PREDMET postojala je jedna tranzitivna ovisnost posredstvom OIB NASTAVNIKA.
- Postupak prevođenja stvara samo jednu novu relaciju.
- U nju je prebačen ovisni atribut, a OIB NASTAVNIKA u njoj postaje ključ.

Prevođenje u treću norm formu (5)

- Nakon prevođenja, obje relacije su u 3NF.
 - U preostaloj verziji PREDMET više nema tranzitivnih ovisnosti pa je zadovoljena definicija 3NF.
 - NASTAVNIK ima samo dva atributa pa je sigurno u 3NF.
- Shematski prikaz prevođenja u 3NF:



Sadržaj Poglavlja 4

4.1. Prva, druga i treća normalna forma

4.2. Boyce-Codd-ova i četvrta normalna forma

4.3. Potreba za normalizacijom

Općenito o višim normalnim formama

- U svojim kasnijim radovima krajem 70-tih, Edgar Codd definirao je pojačanu varijantu 2NF i 3NF koja se zove Boyce-Codd-ova normalna forma.
- Ronald Fagin je u isto vrijeme uveo četvrtu i petu normalnu formu. Postoji i šesta normalna forma.
- U praksi je lako naići na relacije koje odstupaju od 2NF, 3NF.
- Rijetko se susreću relacije u 3NF koje nisu u višim normalnim formama.
- Mi ćemo ipak opisati Boyce-Coddovu i četvrtu normalnu formu.

Prevođenje u BC normalnu formu (1)

- *Determinanta* je atribut (ili kombinacija atributa) unutar neke relacije o kojem je neki drugi atribut unutar iste relacije potpuno funkcionalno ovisan.
- Relacija je u *Boyce-Codd-ovoj normalnoj formi* (oznaka: BCNF) ako je svaka njezina determinanta ujedno i kandidat za ključ.
- Primjer relacije koja nije u BCNF (ona ista koja nije bila u 2NF i 3NF):
UPISAO (JMBAG, ŠIFRA PREDMETA, NASLOV PREDMETA,
OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA, OCJENA) .

Prevođenje u BC normalnu formu (2)

- Ponovo uočavamo funkcionalne ovisnosti:

ŠIFRA PREDMETA → NASLOV PREDMETA ,

ŠIFRA PREDMETA → OIB NASTAVNIKA ,

OIB NASTAVNIKA → BROJ SOBE NASTAVNIKA .

- Postoje dvije determinante:

– To su ŠIFRA PREDMETA i OIB NASTAVNIKA.

– Nijedna od njih nije kandidat za ključ.

Dakle relacija nije u BCNF.

- Relacija koja nije u BCNF smatra se loše oblikovanom, te mogućim izvorom poteškoća. Preporuča se da se ona prevede u BCNF.

Prevođenje u BC normalnu formu (3)

- Postupak prevođenja:
 - Uz polaznu relaciju dodaje se onoliko novih relacija koliko ima različitih determinanti koje nisu kandidat za ključ.
 - Iz polazne relacije *izbacuju se i prebacuju* u nove svi oni atributi koji su ovisni o nekoj tih determinanti.
 - Pritom u jednu novu relaciju idu oni atributi koji su ovisni o istoj determinanti.
 - Uz prebačene attribute, u novu relaciju *prepisuje se* i sama determinanta, te ona postaje ključ u toj novoj relaciji.

Prevođenje u BC normalnu formu (4)

- Ako postupak prevođenja u BCNF primijenimo na polaznu relaciju UPISAO, ona se razbija na tri relacije:

UPISAO (JMBAG, ŠIFRA PREDMETA, OCJENA)

PREDMET (ŠIFRA PREDMETA, NASLOV PREDMETA,
OIB NASTAVNIKA)

NASTAVNIK (OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA) .

- Jednim prevođenjem dobili ono isto rješenje za koje su prije bila potrebna dva prevođenja, najprije u 2NF, zatim u 3NF.

Odnos BCNF prema 2NF i 3NF (1)

- BCNF uključuje u sebi sve zahtjeve koje postavljaju 2NF i 3NF.
 - Relacija koja nije u 2NF ili 3NF ne može biti ni u BCNF.
 - Relacija koja jeste u BCNF mora nužno biti u 2NF i 3NF.
- Je li BCNF zapravo ekvivalentna 2NF i 3NF?
- Ipak nije! BCNF postavlja na relaciju malo jači zahtjev od 2NF i 3NF.

Odnos BCNF prema 2NF i 3NF (2)

- Makar relacija koja je u BCNF nužno mora biti u 2NF i 3NF, može se desiti da relacija koja jest u 2NF i 3NF ipak nije u BCNF.
 - BCNF se može smatrati „trećom-i-pol normalnom formom“.
- Primjeri relacija koje jesu u 2NF i 3NF no nisu u BCNF vrlo su rijetki.
 - Takvi primjeri zasnivaju se na postojanju dva kandidata za ključ.
 - Oba kandidata za ključ su složena i preklapaju se u jednom atributu.

Odnos BCNF prema 2NF i 3NF (3)

- Primjer se odnosi na glazbenu akademiju gdje :
 - Jedan predmet predaje više nastavnika, ali svaki nastavnik predaje samo jedan predmet.
 - Svaki student upisuje više predmeta, no ima samo jednog nastavnika za zadani predmet.
- Odnosi se mogu opisati sljedećom relacijom:
UPISAO (JMBAG, OIB NASTAVNIKA, ŠIFRA PREDMETA) .
- Uz ovako određeni primarni ključ, relacija nije ni u 2NF, jer postoji parcijalna ovisnost
OIB NASTAVNIKA → ŠIFRA PREDMETA .

Odnos BCNF prema 2NF i 3NF (4)

- No mi možemo drukčije izabrati primarni ključ. Ista relacija tada izgleda ovako:
UPISAO (JMBAG, ŠIFRA PREDMETA, OIB NASTAVNIKA) .
- Sad je relacija u 2NF i 3NF, no ne i u BCNF.
 - Naime i dalje postoji ovisnost:
OIB NASTAVNIKA → ŠIFRA PREDMETA .
 - Pritom OIB NASTAVNIKA nije kandidat za ključ.

Odnos BCNF prema 2NF i 3NF (5)

- Na zadnju verziju relacije UPISAO primjenjujemo postupak prevođenja u BCNF:

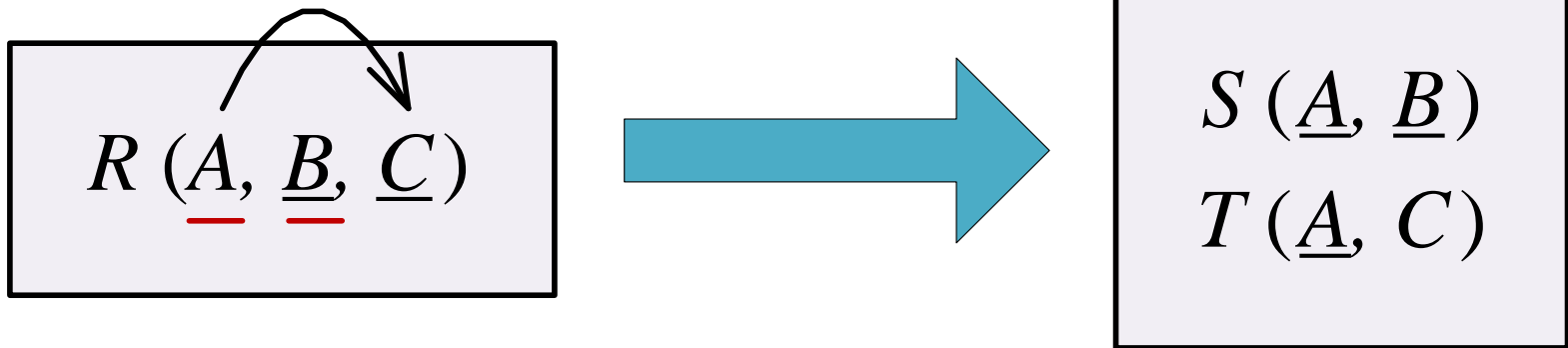
KLASA (JMBAG, OIB NASTAVNIKA)

PREDAJE (OIB NASTAVNIKA, ŠIFRA PREDMETA) .

- Obje novonastale relacije su u BCNF.
 - Iz prve vidimo u čijoj klasi je određen student upisao određeni (to jest nastavnikov) predmet,
 - Iz druge vidimo koji to predmet predaje taj nastavnik.

Odnos BCNF prema 2NF i 3NF (6)

- Shematski prikaz prevođenja u BCNF za slučaj relacije koja jeste u 2NF i 3NF.



- Pretpostavljeno je da polazna relacija ima dva kandidata za ključ,
- Oba su složena i preklapaju se u jednom atributu.

Prevođenje u četvrtu norm formu (1)

- Relacija IZVOZI prikazuje vezu između kompanija, proizvoda i zemalja:

IZVOZI (IME KOMPANIJE, IME PROIZVODA, IME ZEMLJE).

Jedna n -torka izražava činjenicu da zadana kompanija svoj zadani proizvod izvozi u zadanu zemlju.

- Lagano je provjeriti da je relacija u BCNF. U jednom trenutku ona može izgledati ovako.

Prevođenje u četvrtu norm formu (2)

- Izgleda da vrijedi pravilo: *čim kompanija izvozi u neku zemlju, ona odmah izvozi sve svoje proizvode u tu zemlju.*

Ako pravilo vrijedi, tada IZVOZI sadrži redundanciju.

IME KOMPANIJE	IME PROIZVODA	IME ZEMLJE
IBM	Desktop	Francuska
IBM	Desktop	Italija
IBM	Desktop	Velika Britanija
IBM	Mainframe	Francuska
IBM	Mainframe	Italija
IBM	Mainframe	Velika Britanija
HP	Desktop	Francuska
HP	Desktop	Španjolska
HP	Desktop	Irska
HP	Server	Francuska
HP	Server	Španjolska
HP	Server	Irska
Fujitsu	Mainframe	Italija
Fujitsu	Mainframe	Francuska

Prevođenje u četvrtu norm formu (3)

- Redundancija će se eliminirati ako polaznu relaciju IZVOZI zamijenimo s dvije manje:

RADI (IME KOMPANIJE, IME PROIZVODA)

PRODAJE (IME KOMPANIJE, IME ZEMLJE) .

- Odgovarajući podaci:

IME KOMPANIJE	IME PROIZVODA
IBM	Desktop
IBM	Mainframe
HP	Desktop
HP	Server
Fujitsu	Mainframe

IME KOMPANIJE	IME ZEMLJE
IBM	Francuska
IBM	Italija
IBM	Velika Britanija
HP	Francuska
HP	Španjolska
HP	Irska
Fujitsu	Italija
Fujitsu	Francuska

Prevođenje u četvrtu norm formu (4)

- Dosadašnja pravila normalizacije ne pomažu da eliminiramo redundanciju u relaciji IZVOZI.
 - Redundancija nije bila uzrokovana funkcionalnim ovisnostima, već višeznačnim ovisnostima.
- Zadana je relacija s tri atributa: $R(A, B, C)$.
Višeznačna ovisnost od A do B (ozn: $A \twoheadrightarrow B$)
vrijedi ako skup B -vrijednosti koje se u R pojavljuju uz zadani par (A -vrijed, C -vrijed) ovisi samo o A -vrijednosti, a ne i o C -vrijednosti.

Prevođenje u četvrtu norm formu (5)

- Skup proizvoda koje zadana kompanija izvozi u zadanu zemlju ovisi samo o kompaniji a ne o zemlji.
- Skup zemalja u koje kompanija izvozi proizvod ovisi samo o kompaniji a ne o proizvodu.
- Zato vrijede višeznačne ovisnosti:

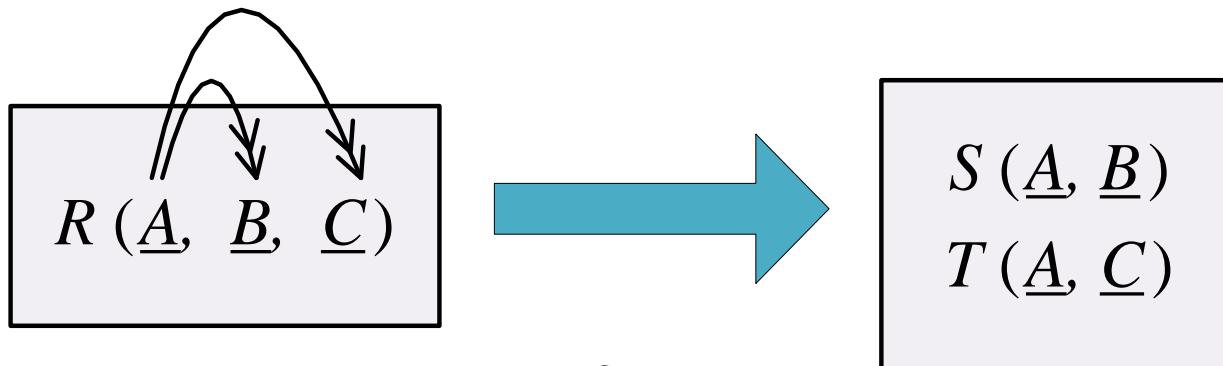
IME KOMPANIJE →→ IME PROIZVODA ,
IME KOMPANIJE →→ IME ZEMLJE .
- Višeznačna ovisnost smatra se nepoželjnom. Da bi se takva ovisnost mogla eliminirati, uvodi se pojam četvrte normalne forme.

Prevođenje u četvrtu norm formu (6)

- Relacija R je u *četvrtoj normalnoj formi* (oznaka: 4NF) ako vrijedi: kad god postoji višeznačna ovisnost u R , na primjer $A \twoheadrightarrow B$, tada su svi atributi od R funkcionalno ovisni o A .
 - R je u 4NF ako je u BCNF i sve višeznačne ovisnosti u R su zapravo funkcionalne ovisnosti.
- U relaciji IZVOZI ni jedna od uočenih višeznačnih ovisnosti nije funkcionalna ovisnost.
 - IZVOZI nije u 4NF i treba je rastaviti na RADI i PRODAJE.

Prevođenje u četvrtu norm formu (7)

- Postupak prevođenja relacije u 4NF:
 - Polaznu relaciju pretvaramo u dvije manje relacije od po dva atributa (odnosno dvije skupine atributa), tako da u njima nema višeznačnih ovisnosti.
 - U svaku od novih relacija stavljamo po dva atributa (odnosno dvije skupine atributa) koji u polaznoj relaciji sudjeluju u istoj višeznačnoj ovisnosti.
- Shematski prikaz prevođenja u 4NF:



Sadržaj Poglavlja 4

4.1. Prva, druga i treća normalna forma

4.2. Boyce-Codd-ova i četvrta normalna forma

4.3. Potreba za normalizacijom

Općenito o potrebi za normalizacijom

- Normalizacija je potrebna zato jer se njome *izbjegavaju teškoće* koje bi nastupile ako bi radili s nenormaliziranim podacima.
- Normalizacija je korisna i zato jer se njome *naknadno otkrivaju i ispravljaju greške* u oblikovanju entiteta, veza i atributa.
- Od normalizacije možemo *odustati samo u nekim rijetkim situacijama*, no i tada moramo biti svjesni eventualnih loših posljedica takve odluke.

Teškoće s nenormaliziranim pod (1)

- Ako relacije nisu normalizirane, dolazi do teškoća kod unosa, promjene i brisanja podataka.
 - Bez obzira o kojoj normalnoj formi je riječ, teškoće su otprilike slične.

- Promatrajmo polaznu verziju relacije UPISAO koja nije u 2NF.

UPISAO (JMBAG, ŠIFRA PREDMETA, NASLOV PREDMETA, OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA, OCJENA) .

- Do teškoća u radu s UPISAO dolazi onda kad preko nje pokušavamo upravljati podacima o nastavnicima ili predmetima. Na primjer:

Teškoće s nenormaliziranim pod (2)

- Ako želimo unijeti novi predmet, to ne možemo učiniti sve dok bar jedan student ne upiše taj predmet.
- Ako želimo promijeniti naslov postojećeg predmeta, tada moramo naći sve n -torke koje sadrže odgovarajuću vrijednost za ŠIFRU PREDMETA, te im promijeniti vrijednost za NASLOV PREDMETA.
 - Bit će onoliko promjena koliko ima studenata koji su upisali predmet.
 - Ako zaboravimo izvršiti neku od promjena, imat ćemo kontradiktorne podatke.
- Pretpostavimo da svi studenti koji su upisali neki predmet naknadno odustanu od tog predmeta. Ako pobrišemo odgovarajuće n -torke, iz baze će nestati svi podaci o tom predmetu.

Teškoće s nenormaliziranim pod (3)

- Promatrajmo polaznu verziju relacije PREDMET koja nije u 3NF.

PREDMET (ŠIFRA PREDMETA, NASLOV PREDMETA,
OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA) .

- Do teškoća u radu s PREDMET dolazi onda kad preko nje pokušavamo mijenjati podacima o nastavnicima. Na primjer:
 - Ne možemo unijeti podatke o novom nastavniku i njegovoj sobi, sve dok ga nismo zadužili s bar jednim predmetom.

Teškoće s nenormaliziranim pod (4)

- Da bismo promijenili broj sobe određenog nastavnika, moramo izvršiti promjenu u svakoj n -torki koja odgovara nekom predmetu kojeg predaje taj nastavnik.
 - Ako zaboravimo neki od predmeta, imat ćemo kontradiktorne podatke.
- Ako nastavnik privremeno ne predaje ni jedan predmet, tada iz baze nestaju svi podaci o njemu i njegovoj sobi.
- Promatramo relaciju UPISAO koja se odnosi na glazbenu akademiju i koja nije u BCNF.
UPISAO (JMBAG, ŠIFRA PREDMETA, OIB NASTAVNIKA) .

Teškoće s nenormaliziranim pod (5)

- Do teškoća u radu UPISAO doći će onda kad preko nje počnemo raditi s podacima o predmetima i nastavnicima. Na primjer:
 - Ne možemo evidentirati činjenicu da zadani nastavnik predaje zadani predmet, sve dok bar jedan student ne upiše taj predmet baš kod tog nastavnika.
 - Veza nastavnika i predmeta zapisana je s velikom redundancijom, onoliko puta koliko ima studenata u klasi tog nastavnika, što otežava ažuriranje.
 - Ako svi studenti u klasi nekog nastavnika odustanu od sudjelovanja u klasi, briše se evidencija da taj nastavnik predaje taj predmet.

Teškoće s nenormaliziranim pod (6)

- Promatrajmo relaciju IZVOZI koja nije u 4NF:
IZVOZI (IME KOMPANIJE, IME PROIZVODA, IME ZEMLJE) .
- Do teškoća u radu s relacijom IZVOZI dolazi zbog toga što ona s velikom redundancijom opisuje dvije nezavisne veze između triju entiteta. Na primjer:
 - Da bismo evidentirali da neka kompanija ima novi proizvod, morat ćemo unijeti onoliko n -torki koliko ima zemalja u koje ta kompanija izvozi svoje proizvode.
 - Da bismo evidentirali da je neka kompanija počela izvoziti u neku novu zemlju, morat ćemo unijeti onoliko n -torki koliko ima proizvoda te kompanije.

Normalizacijom ispravljamo greške (1)

- Nenormalizirani podaci imaju zajedničku osobinu: oni pokušavaju govoriti o više stvari u isto vrijeme.
 - Neki pokušavaju istovremeno opisati više tipova entiteta.
 - Drugi bilježe vezu između entiteta no istovremeno navode i svojstva samih entiteta.
 - Treći opet istovremeno zapisuju više veza.
- To nije u skladu s postupkom oblikovanja opisanim u Poglavljima 2 i 3.
 - Naime, taj postupak, ukoliko je ispravno proveden, morao bi rezultirati relacijama koje govore o jednom i samo jednom entitetu, ili relacijama koje bilježe jednu i samo jednu vezu.

Normalizacijom ispravljamo greške (2)

- Pojavu nenormaliziranih podataka možemo promatrati kao rezultat greške u postupku oblikovanja.
 - Izvor greške obično se nalazi već u oblikovanju konceptualne sheme, dakle u pogrešnom prepoznavanju entiteta, veza i atributa.
- U nastavku detaljnije analiziramo prethodne primjere nenormaliziranih podataka te otkrivamo odgovarajuće greške u konceptualnom oblikovanju.

Normalizacijom ispravljamo greške (3)

- Relacija UPISAO koja nije u 2NF nastala je zato što se događaj upisivanja predmeta od strane studenta pogrešno interpretirao kao zasebni tip entiteta s vlastitim atributima.

UPISAO (JMBAG, ŠIFRA PREDMETA, NASLOV PREDMETA, OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA, OCJENA) .

- Tipovi entiteta zapravo su studenti i predmeti.
- Upisivanje je samo veza između tih tipova.
- Postoji i tip entiteta za nastavnike.

Normalizacijom ispravljamo greške (4)

- Postupak prevođenja najprije u 2NF, zatim u 3NF, postepeno ispravlja te greške.

UPISAO (JMBAG, ŠIFRA PREDMETA, OCJENA)

PREDMET (ŠIFRA PREDMETA, NASLOV PREDMETA,
OIB NASTAVNIKA)

NASTAVNIK (OIB NASTAVNIKA, BROJ SOBE NASTAVNIKA) .

- UPISAO se sad reducira u oblik koji služi isključivo za bilježenje veze između studenata i predmeta.
- Nastaju nove relacije PREDMET i NASTAVNIK koje odgovaraju istoimenim entitetima.
- Veza između predmeta i nastavnika ispravno se realizira preko stranog ključa u relaciji PREDMET.

Normalizacijom ispravljamo greške (5)

- Relacija UPISAO za glazbenu akademiju koja jeste u 2NF i 3NF no nije u BCNF, nastala je zato što smo odnose između nastavnika, predmeta i studenata pogrešno tumačili kao ternarnu vezu.

UPISAO (JMBAG, ŠIFRA PREDMETA, OIB NASTAVNIKA) .

- Zapravo se radilo o dvije nezavisne binarne veze.
- Da smo to odmah uočili, podaci bi odmah bili prikazani relacijama koje smo dobili prevođenjem u BCNF:

KLASA (JMBAG, OIB NASTAVNIKA)

PREDAJE (OIB NASTAVNIKA, ŠIFRA PREDMETA) .

Normalizacijom ispravljamo greške (6)

- Relacija IZVOZI koja nije u 4NF nastala je zato što smo odnose između kompanija, proizvoda i zemalja pogrešno prikazali jednom ternarnom vezom.

IZVOZI (IME KOMPANIJE, IME PROIZVODA, IME ZEMLJE) .

- Zapravo je riječ o dvjema binarnim vezama.
- Da smo krenuli od tih binarnih veza, odmah bi nastale one relacije koje smo dobili prevođenjem u 4NF.

RADI (IME KOMPANIJE, IME PROIZVODA)

PRODAJE (IME KOMPANIJE, IME ZEMLJE) .

Normalizacijom ispravljamo greške (7)

- Pravila normalizacije zapravo su formalni opis intuitivno prihvatljivih principa o zdravom i prirodnom oblikovanju entiteta, veza i atributa.
- Ako oblikovanje uspijemo provesti na ispravan način, odmah ćemo dobiti relacije u visokim normalnim formama i normalizacija neće imati što raditi.
- To naravno ne znači da je postupak normalizacije nepotreban, baš naprotiv.

Normalizacijom ispravljamo greške (8)

- Naime, greške su uvijek moguće i mi nikad ne znamo je li do njih došlo ili nije.
- Normalizacija predstavlja mehanizam naknadnog prepoznavanja i ispravljanja grešaka.
- Svođenjem u normalnu formu pogrešno oblikovane relacije ponovo se vraćaju u onaj oblik koje bi imale da grešaka nije bilo.
- Zahvaljujući normalizaciji, postupak oblikovanja postaje samokorigirajući proces, gdje se greške počinjene u ranijim fazama uspješno ispravljaju u kasnijim fazama.

Kad se može odustati od normaliz (1)

- Postoje razlozi zbog kojih iznimno možemo odustati od normalizacije. Navodimo dva razloga.

- **Složeni atribut.** Nekoliko atributa u relaciji čine cjelinu koja se u aplikacijama nikad ne rastavlja na sastavne dijelove. Na primjer:

KUPAC (OIB KUPCA, PREZIME, IME, POŠTANSKI BROJ, IME GRADA, ULICA I KUĆNI BROJ) .

- Strogo govoreći, IME GRADA je funkcionalno ovisno o POŠTANSKOM BROJU, pa relacija nije u 3NF.
 - No mi znamo da POŠTANSKI BROJ, IME GRADA, te ULICA I KUĆNI BROJ čine cjelinu koja se zove adresa.
 - Podaci iz adrese rabe i ažuriraju „u paketu“.
 - Nije preporučljivo ovu relaciju razbijati na dvije.

Kad se može odustati od normaliz (2)

- **Efikasna uporaba podataka.** Normalizacijom se velike relacije razbijaju na mnogo manjih. U aplikacijama je često potrebno podatke iz malih relacija ponovo sastavljati u veće nenormalizirane n -torke. Ako imamo aplikacije kod kojih se zahtijeva izuzetno velika brzina odziva, tada je bolje da su podaci za njih već pripremljeni u nenormaliziranom obliku.
- Projektant baze podataka treba procijeniti kada treba provesti normalizaciju do kraja a kada ne.
 - Za tu procjenu je važno razumijevanje značenja podataka i načina kako će se oni zaista rabiti.