



Normalizacija baze podataka

- Normalizacijom se nastoji otkloniti potreba za višestrukim zapisivanjem istih podataka.

- Normalne forme su temeljno formalno mjerilo valjanosti relacijske BP

- **Normalizacija** predstavlja primjenu:

matematičkih

i formalnih pravila

- osigurava se **ispravno postavljanje modela** podataka i njihova **logička povezanost**.



Normalizacija baze podataka

NENORMALIZIRANA TABLICA

ŠifraStudenta	ŠifrePredmeta
1	10,14,16
2	14,15,16
3	11
4	10, 140
5	16,10

ŠifraStudenta	ŠifraPredmeta1	ŠifraPredmeta2	ŠifraPredmeta3
1	10	14	16
2	14	15	16
3	11		
4	10	140	
5	16	10	

ŠifraStudenta	ŠifraPredmeta
1	10
2	14
3	11
4	10
5	16
1	14
1	16
2	15
2	16
4	140
5	10



Normalizacija baze podataka

FUNKCIJSKA OVISNOST

K	L	M
1	Da	Jučer
2	Ne	Danas
3	Možda	Sutra
4	Skoro	Jučer
5	Možda	Sutra

U tablici T kojoj se nalaze atributi A i B (koji mogu biti složeni), atribut B funkcijski ovisi o atributu A ako za svaku vrijednost atributa A postoji samo jedna vrijednost atributa B.

Može se reći da atribut A određuje atribut B, a simbolično se to označava $A \rightarrow B$.

Da bi vrijedila funkcijska ovisnost $A \rightarrow B$, u tablici ne smiju postojati dva reda s istom vrijednošću atributa A, a različitim vrijednostima atributa B.




Normalizacija baze podataka

1. NORMALNA FORMA


Tablica se nalazi u 1. normalnoj formi ako su svi neključni atributi funkcijski ovisni o primarnom ključu.

- pravilo je jednostavno i proizlazi iz definicije primarnog ključa
- naglašava transformacijsko pravilo prikaza viševrijednosnih atributa.



Funkcijska zavisnost atributa

A	B	C
a	100	6
b	200	7
c	300	8
d	200	9



Funkcijska zavisnost atributa

A	B	C
☹	3	KL
✓	0	TR
☒	7	MN
✓	2	TR



Normalizacija baze podataka

Primjer:

StudentID	Ime	Prezime	DatumRođenja	Telefon	Hobi
1	Ratimir	Mirić	5.11.1988	093/741-258	
2	Izabela	Izić	11.2.1992	093/852-456	ples
3	Zvezdoslav	Starić	29.6.1986	093/753-159	astronomija
4	Marinela	Orlović	30.7.1990	093/963-147	planinarenje, vožnja biciklom

- Prema definiciji, da bi tablica bila u prvoj normalnoj formi, svi neključni atributi moraju funkcijski ovisiti o primarnom ključu (StudentID).

- Hobi ne ovisi o StudentID jer za isti StudentID (4) postoje dva različita hobija (planinarenje i vožnja biciklom).

Rješenje:

StudentID	Ime	Prezime	DatumRođenja	Telefon
1	Ratimir	Mirić	5.11.1988	093/741-258
2	Izabela	Izić	11.2.1992	093/852-456
3	Zvezdoslav	Starić	29.6.1986	093/753-159
4	Marinela	Orlović	30.7.1990	093/963-147

StudentID	Hobi
2	ples
3	astronomija
4	planinarenje
4	vožnja biciklom



Normalizacija baze podataka – 2NF

StudentID	Datum	PredmetID	NazivPredmeta	ECTSPredmeta	Ocjena
5	10.9.2011	20	Operacijski sustavi	5	3
6	23.6.2011	28	Informacijski sustavi	6	5
6	23.6.2011	25	Baze podataka	6	2
4	10.7.2011	30	Računalne mreže	5	1
5	27.9.2011	30	Računalne mreže	5	3
6	27.9.2011	30	Računalne mreže	5	5

- Tablica ISPIT ima samo jednog kandidata za ključ koji se sastoji od atributa StudentID, Datum i PredmetID.

- Tablica **se nalazi u 1NF** jer se primarni ključ (kombinacija tri atributa) neće ponavljati (isti student u istom danu može najviše jednom polagati jedan predmet).

- Anomalije :

- Anomalija unosa: ne može se unijeti podatak o nazivu predmeta i o ECTS bodovima koje on nosi dokle god netko ne izađe na ispit iz tog predmeta. Također, kod svakog polaganja ispita iz npr. Računalnih mreža potrebno je ponovno unositi sve podatke o tom predmetu.
- Anomalija izmjene: ako se promijeni npr. naziv predmeta potrebno je izvršiti izmjene onoliko puta koliko je taj predmet bio polagan.
- Anomalija brisanja: brisanjem zadnjeg ispita iz nekog predmeta gube se i svi podaci o predmetu.



Normalizacija baze podataka – 2NF

- Definicija: Tablica se nalazi u drugoj normalnoj formi ako **se nalazi u 1NF i ako je svaki neključni atribut potpuno funkcijski ovisan o kandidatima ključa**.
- U tablici T u kojoj se nalaze atributi A i B (koji mogu biti složeni), atribut B **potpuno funkcijski ovisi** o atributu A ako B funkcijski ovisi o A i ne ovisi funkcijski niti o jednom manjem podskupu od A.
- To znači da svi neključni atributi moraju biti funkcijski ovisni o kandidatu ključa, ali ne i o dijelu ključa.
- Pravilo druge normalne forme se stoga tiče onih tablica koje imaju složene kandidate ključa.
- Ako se tablica nalazi u 1NF, a kandidat ključa se sastoji od samo jednog atributa, onda je tablica automatski u drugoj normalnoj formi.



Normalizacija baze podataka – 2NF

- Rješenje primjera je da se stupce koji funkcijski ovise o dijelu primarnog ključa (to su stupci NazivPredmeta i ECTSPredmeta) premjeste u novu tablicu, a kako se ne bi izgubili odnosi među podacima, u novu tablicu ulazi i stupac koji je dio primarnog ključa o kojem oni ovise (PredmetID).
- PredmetID će postati primarni ključ nove tablice PREDMET.

StudentID	Datum	PredmetID	Ocjena
4	10.7.2011	30	1
5	10.9.2011	20	3
5	27.9.2011	30	3
6	23.6.2011	25	2
6	23.6.2011	28	5
6	27.9.2011	30	5

PredmetID	NazivPredmeta	ECTSPredmeta
20	Operacijski sustavi	5
25	Baze podataka	6
28	Informacijski sustavi	6
30	Računalne mreže	5

Normalizacija baze podataka – 3NF

Tablica se nalazi u trećoj normalnoj formi ako se nalazi u 2NF i ako ne postoji tranzitivna funkcijska ovisnost nijednog neključnog atributa o kandidatima ključa.

U tablici T u kojoj se nalaze atributi A, B i C, ako vrijedi $A \rightarrow B$, B ne ovisi o A, i $B \rightarrow C$, tada C tranzitivno funkcijski ovisi o atributu A, što se zapisuje $A \rightarrow B \rightarrow C$.

PredmetID	NazivPredmeta	ECTSPredmeta	PredavačID	Ime	Prezime
20	Operacijski sustavi	5	104	Viktor	Marohnić
25	Baze podataka	6	108	Ivan	Britvić
28	Informacijski sustavi	6	111	Ivan	Klasić
30	Računalne mreže	5	105	Damir	Delija
31	Multimedijske mreže i sustavi	5	105	Damir	Delija

Anomalije koje se javljaju u ovakvoj tablici:

- Anomalija unosa: ne mogu se unijeti podaci o predavaču koji ne predaje niti jedan predmet. Također, za svaki predmet kojeg predaje neki predavač potrebno je ponovno unositi sve podatke o tom predavaču.
- Anomalija izmjene: ako se promijeni npr. prezime predavača (češće predavačice), potrebno je izvršiti izmjene na svakom predmetu kojeg predavač(ica) predaje.
- Anomalija brisanja: brisanjem zadnjeg predmeta kojeg predaje neki predavač, gube se i svi podaci o tom predavaču.

Normalizacija baze podataka – 3NF

PredmetID	NazivPredmeta	ECTSPredmeta	PredavačID	Ime	Prezime
20	Operacijski sustavi	5	104	Viktor	Marohnić
25	Baze podataka	6	108	Ivan	Britvić
28	Informacijski sustavi	6	111	Ivan	Klasić
30	Računalne mreže	5	105	Damir	Delija
31	Multimedijske mreže i sustavi	5	105	Damir	Delija

Rješenje je da se stupce koji tranzitivno funkcijski ovise o primarnom ključu (to su stupci Ime i Prezime) premjeste u novu tablicu PREDAVAČ, u koju ulazi i stupac koji posreduje u tranzitivnoj vezi (PredavačID). PredavačID će postati primarni ključ nove tablice PREDAVAČ.

PredmetID	NazivPredmeta	ECTSPredmeta	PredavačID
20	Operacijski sustavi	5	104
25	Baze podataka	6	108
28	Informacijski sustavi	6	111
30	Računalne mreže	5	105
31	Multimedijske mreže i sustavi	5	105

PredavačID	Ime	Prezime
104	Viktor	Marohnić
105	Damir	Delija
108	Ivan	Britvić
111	Ivan	Klasić



Normalizacija baze podataka

BOYCE-CODDOVA NORMALNA FORMA (BCNF)

- Boyce-Coddova normalna forma obrađuje poseban slučaj kad tablica ima više kandidata za ključ koji se međusobno djelomično preklapaju.
- Da bi Boyce-Coddova normalna forma došla u razmatranje, potrebno je da budu ispunjeni slijedeći uvjeti :
 - Tablica mora imati dva ili više kandidata za primarni ključ.
 - Najmanje dva od tih kandidata moraju biti složeni.
 - Kandidati za ključeve moraju imati neke zajedničke atribute.

Definicija.

Tablica se nalazi u Boyce-Coddovoj normalnoj formi ako su sve njezine netrivialne funkcijske ovisnosti ujedno i funkcijske ovisnosti o ključu.

- Funkcijska ovisnost $A \rightarrow B$ je trivijalna ako je B podskup od A.



Normalizacija baze podataka

BOYCE-CODDOVA NORMALNA FORMA (BCNF)

U tablici SLUŠA koja opisuje odnose studenata, predmeta i predavača vrijedi da:

- jedan student sluša predmet kod jednog predavača (dva predavača ne mogu nekom studentu predavati isti predmet)
- jedan predavač predaje samo jedan predmet (jedan predavač ne može predavati više predmeta)
- jedan predmet može predavati više predavača

Student	Predmet	Predavač
Ivica	Elektrotehnika	Divić
Marica	Matematika	Čančarević
Đuro	Programiranje	Bosnić
Barica	Programiranje	Britvić

Normalizacija baze podataka

BOYCE-CODDOVA NORMALNA FORMA (BCNF)

- Ovakva tablica ima dva kandidata za ključ: (Student, Predmet) i (Student, Predavač).
- Kandidati imaju jedan zajednički atribut: Student.
- Ako se za ključ odabere (Student, Predavač), tablica ne zadovoljava 2NF
- Odabirom (Student, Predmet) za primarni ključ relacija je u 3NF.
- Međutim, tablica nije u BCNF jer osim funkcijske ovisnosti (Student, Predmet) → Predavač, postoji i Predavač → Predmet, dakle funkcijska ovisnost o neključnom atributu.
- Ovdje se zapravo radi o tranzitivnoj ovisnosti:
(Student, Predmet) → Predavač → Predmet
- Ovakva tranzitivna ovisnost ključnog atributa preko neključnog atributa može postojati u 3NF, no nije dopuštena u BCNF.

Student	Predmet	Predavač
Ivica	Elektrotehnika	Divić
Marica	Matematika	Čančarević
Đuro	Programiranje	Bosnić
Barica	Programiranje	Britvić

Normalizacija baze podataka

BOYCE-CODDOVA NORMALNA FORMA (BCNF)

- Anomalije koje se javljaju u ovakvoj tablici:
 - Anomalija unosa: ne može se unesti podatak o tome koji predavač predaje neki predmet sve dok barem jedan student ne upiše taj predmet kod tog predavača.
 - Anomalija promjene: ako se promijeni predavač na nekom predmetu, potrebno je izvršiti izmjene onoliko puta koliko studenata sluša taj predmet.
 - Anomalija brisanja: brisanjem retka sa zadnjim studentom koji sluša predmet kod nekog predavača, briše se i podatak o tome koji predmet taj predavač predaje

Student	Predmet	Predavač
Ivica	Elektrotehnika	Divić
Marica	Matematika	Čančarević
Đuro	Programiranje	Bosnić
Barica	Programiranje	Britvić

Normalizacija baze podataka

BOYCE-CODDOVA NORMALNA FORMA (BCNF)

Rješenje

- Rješenje (dovođenje u BCNF) je da se neključni atribut koji posreduje u tranzitivnoj funkcijskoj ovisnosti (Predavač) komponente ključa (Predmet) izdvoji u novu tablicu skupa s atributom koji o njemu ovisi (Predmet).

Student	Predmet	Predavač
Ivica	Elektrotehnika	Divić
Marica	Matematika	Čančarević
Đuro	Programiranje	Bosnić
Barica	Programiranje	Britvić

Student	Predavač
Ivica	Divić
Marica	Čančarević
Đuro	Bosnić
Barica	Britvić

Predavač	Predmet
Divić	Elektrotehnika
Čančarević	Matematika
Bosnić	Programiranje
Britvić	Programiranje

Normalizacija baze podataka 4NF

Primjer: Svaki predmet predaje više predavača te se za svaki predmet koristi više udžbenika.

Tablica se nalazi u 3NF, no postoji evidentan problem dodavanja novog predavača za postojeći predmet za koji se koristi više udžbenika.

Primjerice, kad bi se trebalo dodati novog predavača za predmet Informacijski sustavi, potrebno bi bilo dodati tri reda u tablicu – po jedan za svaki udžbenik koji se koristi za taj predmet.

Predmet	Predavač	Udžbenik
Informacijski sustavi	Klasić	Uvod u informacijske sustave
Informacijski sustavi	Klasić	Projektiranje IS-a
Informacijski sustavi	Klasić	Poslovni IS-i
Informacijski sustavi	Galinec	Uvod u informacijske sustave
Informacijski sustavi	Galinec	Projektiranje IS-a
Informacijski sustavi	Galinec	Poslovni IS-i
Objektno programiranje	Čaklović	Demistificirani C++
Objektno programiranje	Čaklović	Thinking in C++
Objektno programiranje	Balen	Demistificirani C++
Objektno programiranje	Balen	Thinking in C++

Ovdje se pojavljuje višeznačna ovisnost: s pojedinim predmetom je povezan skup predavača (a ne samo jedan), te isto tako s predmetom je povezan i skup udžbenika



Normalizacija baze podataka 4NF

Tablica se nalazi u četvrtoj normalnoj formi ako i samo ako su sve njezine netrivialne višeznačne ovisnosti ujedno funkcijske ovisnosti od kandidata ključa.

Četvrta normalna forma uvodi pojam višeznačne ovisnosti koja se može definirati na sljedeći način:

U relaciji R s atributima A, B, C vrijedi da B višeznačno ovisi o A ($A \twoheadrightarrow B$) ako i samo ako skup vrijednosti atributa iz B za dani par vrijednosti atributa iz A i C ovisi isključivo o vrijednostima atributa iz A, a neovisan je od vrijednosti atributa iz C.

Višeznačna ovisnost $A \twoheadrightarrow B$ je trivijalna (na relaciji R) ako je B podskup od A ili ako je R binarna relacija.

U primjeru vrijedi Predmet \twoheadrightarrow Udžbenik budući da se udžbenik koristi za točno određeni predmet, a ne ovisi o predavaču koji taj predmet predaje.

Višeznačne ovisnosti dolaze u paru, pa osim što vrijedi Predmet \twoheadrightarrow Udžbenik, vrijedi i Predmet \twoheadrightarrow Predavač. Relacija iz primjera ima, dakle, dvije višeznačne ovisnosti koje nisu funkcijske ovisnosti (pa niti funkcijske ovisnosti od kandidata ključa) zbog čega ona nije u 4NF.



Normalizacija baze podataka 4NF

Shemu relacije R na kojoj postoji prave višeznačne ovisnosti (prava višeznačna ovisnost je ona netrivialna višeznačna ovisnost koja nije funkcijska ovisnost) oblika:

$A \twoheadrightarrow B$ i $A \twoheadrightarrow C$

dekomponira se na sheme relacija:

$R_1(A, B)$ i $R_2(A, C)$

Nakon što se tablica rastavi na dvije, obje nove tablice će biti u 4NF. Vrijedi pravilo da su sve tablice s dva atributa uvijek u četvrtoj normalnoj formi.

Predmet	Predavač	Udžbenik
Informacijski sustavi	Klasić	Uvod u informacijske sustave
Informacijski sustavi	Klasić	Projektiranje IS-a
Informacijski sustavi	Klasić	Poslovni IS-i
Informacijski sustavi	Galinec	Uvod u informacijske sustave
Informacijski sustavi	Galinec	Projektiranje IS-a
Informacijski sustavi	Galinec	Poslovni IS-i
Objektno programiranje	Čaklović	Demistificirani C++
Objektno programiranje	Čaklović	Thinking in C++
Objektno programiranje	Balen	Demistificirani C++
Objektno programiranje	Balen	Thinking in C++

Predmet	Predavač
Informacijski sustavi	Klasić
Informacijski sustavi	Galinec
Objektno programiranje	Čaklović
Objektno programiranje	Balen

Predmet	Udžbenik
Informacijski sustavi	Uvod u informacijske sustave
Informacijski sustavi	Projektiranje IS-a
Informacijski sustavi	Poslovni IS-i
Objektno programiranje	Demistificirani C++
Objektno programiranje	Thinking in C++



Normalizacija baze podataka

Pri **dekompoziciji tablica**, mora vrijediti da se **postupkom razdvajanja tablica ne izgubi** niti jedna informacija, niti se pojave informacije koje ne postoje u polaznoj tablici.

Postoje dva pravila:

- 1) Svaka funkcijska zavisnost tablice **T** može biti logički izvedena iz funkcijskih zavisnosti u tablicama **R** i **S**, koje nastanu dekompozicijom tablice T.
- 2) Zajednički atribut tablica **R** i **S** mora biti ključ u barem jednoj od tih tablica.



Normalizacija baze podataka

Dekomponiranjem se model podataka može uvijek dovesti u 4NF bez gubitka informacija, ali ne uvijek i bez gubitka višeznačnih ovisnosti.

5NF.

Relacija je u 5NF ako je u 4NF, a ne vidi se nikakva mogućnost za njeno daljnje razložno dekomponiranje.

Viša NF \Rightarrow veći broj tablica \Rightarrow veći broj čitanja diska

Razlozi za dovođenje u (barem) 3NF:

1. Redundantni podaci nepotrebno zauzimaju prostor
2. Anomalija održavanja
 - a) Nemogućnost upisa
 - b) Otežano mijenjanje sadržaja
 - c) Brisanje neki podataka može dovesti do gubitka drugih podataka