

Normalizacija podatkovnih baz

Funkcionalne odvisnosti

Odvisnosti med atributi relacijske sheme so sredstvo s katerim povemo, katere vrednosti v relaciji so možne in katere ne morejo obstajati. Poznamo funkcionalne, večvrednostne in stične odvisnosti. **Funkcionalne odvisnosti** predstavljajo razmerje med atributi relacijske sheme, ki omejuje dopustne vrednosti atributov v n -tericah relacije.

Definicija. Naj bosta X in Y neprazni podmnožici atributov relacijske sheme R , $X, Y \subseteq R$. Atributi X *funkcionalno določajo* attribute Y kar označimo $X \rightarrow Y$, če v nobeni relaciji s shemo R ne moreta obstajati dve n -terici, ki bi se ujemali v vrednostih atributov X in se ne bi ujemali v vrednostih atributov Y . Formalno,

$$X \rightarrow Y \in F_R \Leftrightarrow \forall r(Sh(r) = R \Rightarrow \forall t_1, t_2(t_1, t_2 \in r \Rightarrow (t_1.X = t_2.X \Rightarrow t_1.Y = t_2.Y))),$$

kjer je F_R množica funkcionalnih odvisnosti, ki veljajo med atributi sheme R . Levi strani funkcionalne odvisnosti X pravimo **determinanta**.

Identifikacija funkcionalnih odvisnosti

Funkcionalne odvisnosti lahko identificiramo na podlagi:

- razumevanja relacijske sheme (zdrav razum, dokumentacija, opis),
- reprezentativne množice podatkov (odvisnosti morajo veljati ves čas).

Pri tem želimo, da so funkcionalne odvisnosti **popolne**, kar pomeni, da so vse determinante minimalne. Zaradi enostavnosti navadno poiščemo **minimalno pokritje** množice funkcionalnih odvisnosti.

1. naloga. Naj bo *ZAPOSLENI* relacijska shema relacije *zaposleni*. Identificirajte funkcionalne odvisnosti na podlagi razumevanja relacijske sheme.

ID	ImePriimek	Položaj	Placa	PodružnicaID	PodružnicaNaslov
7	Luka Ribič	direktor	75000	P001	Kongresni trg 12, Ljubljana
42	Matej Novak	programer	24000	P013	Večna Pot 113, Ljubljana

Vsak zaposleni ima enolično številko zaposlenega. Vsaka podružnica se nahaja na natanko eni lokaciji in

na vsaki lokaciji je največ ena podružnica. Plača zaposlenega je odvisna od položaja zaposlenega ter od podružnice, v kateri je zaposlen.

$$F_{ZAPOsLENI}^{min} = \{ID \rightarrow \{ImePriimek, Polozaj, PodruznicaID\}, PodruznicaID \rightarrow PodruznicaNaslov, PodruznicaNaslov \rightarrow PodruznicaID, \{Polozaj, PodruznicaID\} \rightarrow Placa\}$$

2. naloga. Naj bo $R = ABCDE$ relacijska shema relacije r . Identificirajte funkcionalne odvisnosti na podlagi reprezentativne množice podatkov.

A	B	C	D	E
a	b	z	w	a
e	b	r	w	b
a	d	z	w	c
e	d	r	w	d
a	f	z	s	e
e	f	r	s	f

$$F_R^{min} = \{A \rightarrow C, B \rightarrow D, C \rightarrow A, E \rightarrow ABCD, AB \rightarrow E, BC \rightarrow E\}$$

Izpeljava funkcionalnih odvisnosti

Definicija. Naj bosta X in Y podmnožici atributov relacijske sheme R , $X, Y \subseteq R$. Množica funkcionalnih odvisnosti F_R logično implicira odvisnost $X \rightarrow Y$, če vsaka relacija s shemo R , ki tako zadošča odvisnostim F_R , zadošča tudi odvisnosti $X \rightarrow Y$. Označimo,

$$F_R \models X \rightarrow Y.$$

Armstrongovi aksiomi

- **A1:** refleksivnost $Y \subseteq X \Rightarrow X \rightarrow Y$
- **A2:** razširitev $\{X \rightarrow Y\} \models XZ \rightarrow YZ$
- **A3:** tranzitivnost $\{X \rightarrow Y, Y \rightarrow Z\} \models X \rightarrow Z$

Armstrongovi aksiomi (oziroma izreki) so smiselni, ker generirajo samo veljavne odvisnosti, in popolni, ker generirajo vse veljavne odvisnosti.

Inferenčna pravila

- **I1:** dekompozicija $\{X \rightarrow YZ\} \models X \rightarrow Y$

$$A1 : Y \subseteq YZ \Rightarrow YZ \rightarrow Y$$

$$A3 : \{X \rightarrow YZ, YZ \rightarrow Y\} \models X \rightarrow Y$$

- **I2:** združitev $\{X \rightarrow Y, X \rightarrow Z\} \models X \rightarrow YZ$

$$A2 : \{X \rightarrow Y\} \models X \rightarrow YX$$

$$A2 : \{X \rightarrow Z\} \models YX \rightarrow YZ$$

$$A3 : \{X \rightarrow YX, YX \rightarrow YZ\} \models X \rightarrow YZ$$

- **I3:** psevdotranzitivnost $\{X \rightarrow Y, WY \rightarrow Z\} \models WX \rightarrow Z$

$$A2 : \{X \rightarrow Y\} \models WX \rightarrow WY$$

$$A3 : \{WX \rightarrow WY, WY \rightarrow Z\} \models WX \rightarrow Z$$

3. naloga. Naj bo $R = ABCDXYZW$ relacijska shema ter F_R in E_R pripadajoči množici funkcionalnih odvisnosti. Ali množica odvisnosti F_R logično implicira vse odvisnosti v množici E_R ? Odgovor utemeljite le z uporabo Armstrongovih aksiomov.

$$F_R = \{A \rightarrow BC, B \rightarrow D, C \rightarrow D, X \rightarrow ZW, X \rightarrow Y, YW \rightarrow Z\}$$

$$E_R = \{A \rightarrow BCD, B \rightarrow D, AC \rightarrow BD, X \rightarrow W\}$$

Množica F_R logično implicira množico E_R .

$$A \rightarrow BCD$$

$$A2 : \{C \rightarrow D\} \models BC \rightarrow BCD$$

$$A3 : \{A \rightarrow BC, BC \rightarrow BCD\} \models A \rightarrow BCD$$

$$B \rightarrow D$$

$$AC \rightarrow BD$$

$$A2 : \{A \rightarrow BC\} \models AC \rightarrow BC$$

$$A2 : \{C \rightarrow D\} \models BC \rightarrow BD$$

$$A3 : \{AC \rightarrow BC, BC \rightarrow BD\} \models AC \rightarrow BD$$

$$X \rightarrow W$$

$$A1 : W \subseteq ZW \Rightarrow ZW \rightarrow W$$

$$A3 : \{X \rightarrow ZW, ZW \rightarrow W\} \models X \rightarrow W$$

4. naloga. Naj bo $R = ABCDEFGH$ relacijska shema ter F_R pripadajoča množica funkcionalnih odvisnosti. Ali množica odvisnosti F_R logično implicira podano odvisnost f ? Odgovor utemeljite z uporabo Armstrongovih aksiomov ali inferenčnih pravil.

$$F_R = \{AB \rightarrow D, AC \rightarrow FG, B \rightarrow G, D \rightarrow B\}$$

4.a) $f = AC \rightarrow C$

$$A1 : C \subseteq AC \Rightarrow AC \rightarrow C$$

4.b) $f = AC \rightarrow G$

$$I1 : \{AC \rightarrow FG\} \models AC \rightarrow G$$

4.c) $f = H \rightarrow DE$

$$\text{Ne, ker je } (H)_{F_R}^+ = H.$$

4.d) $f = AB \rightarrow DG$

$$A2 : \{AB \rightarrow D\} \models AB \rightarrow BD$$

$$A2 : \{B \rightarrow G\} \models BD \rightarrow DG$$

$$A3 : \{AB \rightarrow BD, BD \rightarrow DG\} \models AB \rightarrow DG$$

4.e) $f = AD \rightarrow BG$

$$A3 : \{D \rightarrow B, B \rightarrow G\} \models D \rightarrow G$$

$$I2 : \{D \rightarrow B, D \rightarrow G\} \models D \rightarrow BG$$

$$A2 : D \rightarrow BG \models AD \rightarrow ABG$$

$$I1 : AD \rightarrow ABG \models AD \rightarrow BG$$

4.f) $f = AEC \rightarrow BH$

$$\text{Ne, ker je } (AEC)_{F_R}^+ = AECFG.$$

Zaprtje množice atributov

Definicija. Naj bo F_R množica funkcionalnih odvisnosti relacijske sheme R in X podmnožica atributov

sheme R , $X \subseteq R$. **Zaprtje množice atributov** X glede na množico odvisnosti F_R je množica vseh atributov sheme R , ki so funkcionalno odvisni od X . Formalno,

$$X_{F_R}^+ = \{Y \mid Y \in R \wedge F_R \models X \rightarrow Y\}.$$

Po inferenčnem pravilu I2 nemudoma sledi $X \rightarrow X_{F_R}^+$. Za izračun zaprtja $X_{F_R}^+$ lahko uporabimo naslednji naiven algoritem.

```

INPUT:  $X, F$ 
OUTPUT:  $X_F^+$ 

 $X_F^+ = X;$ 
DO
  FOR EACH  $Y \rightarrow Z \in F$  DO
    IF  $Y \subseteq X_F^+$  THEN
       $X_F^+ = X_F^+ \cup Z;$ 
WHILE  $X_F^+$  has changed;
  
```

Primer. Naj bo $R = ABCDEFG$ relacijska shema in F_R pripadajoča množica funkcionalnih odvisnosti. Poiščite zaprtje množice atributov $X = EF$.

$$F_R = \{A \rightarrow B, BE \rightarrow G, EF \rightarrow A, D \rightarrow AC\}$$

Z uporabo algoritma dobimo $EF \mapsto AEF \mapsto ABEF \mapsto ABEFG$ in tako $X_{F_R}^+ = ABEFG$.

5. naloga. Naj bo $R = ABCDEF$ relacijska shema in F_R pripadajoča množica funkcionalnih odvisnosti. Poiščite zaprtje množice atributov X .

$$F_R = \{A \rightarrow CDE, BE \rightarrow CF, C \rightarrow E, D \rightarrow B, E \rightarrow BCD\}$$

5.a) $X = ABCDEF$

$$(ABCDEF)_{F_R}^+ = R$$

5.b) $X = BF$

$$(BF)_{F_R}^+ = BF$$

5.c) $X = E$

$$(E)_{F_R}^+ = BCDEF$$

5.d) $X = AD$

$$(AD)_{F_R}^+ = R$$

5.e) $X = ABD$

$$(ABD)_{F_R}^+ = R$$

6. **naloga.** Naj bo $R = ABCDE$ relacijska shema in F_R pripadajoča množica funkcionalnih odvisnosti. Katere izmed naslednjih odvisnosti ni moč implicirati z množico odvisnosti F_R ?

$$F_R = \{AB \rightarrow CE, B \rightarrow D, D \rightarrow A\}$$

- $E \rightarrow E$,
- $BC \rightarrow DE$,
- $AD \rightarrow CE$,
- $B \rightarrow E$,
- $AB \rightarrow A$ ali
- nobene izmed naštetih.

$$AD \rightarrow CE, \text{ ker velja } CE \not\subseteq (AD)_{F_R}^+.$$

7. **naloga.** Naj bo $R = ABCDEF$ relacijska shema in F_R pripadajoča množica funkcionalnih odvisnosti. Koliko odvisnosti, ki veljajo za shemo R , ima determinanto enako X ?

$$F_R = \{AB \rightarrow CE, B \rightarrow D, D \rightarrow A, A \rightarrow F\}$$

7.a) $X = AF$

$$2^2 - 1 = 3, \text{ ker je } (AF)_{F_R}^+ = AF.$$

7.b) $X = BC$

$$2^6 - 1 = 63, \text{ ker je } (BC)_{F_R}^+ = ABCDEF.$$

7.c) $X = F$

$$2^1 - 1 = 1, \text{ ker je } (F)_{F_R}^+ = F.$$

Identifikacija ključev

Definicija. Naj bo F_R množica funkcionalnih odvisnosti relacijske sheme R in X podmnožica atributov sheme R , $X \subseteq R$. Atributi X so **kandidat za ključ** sheme R in pripadajočih relacij, če velja:

- atributi X funkcionalno določajo vse attribute sheme R ,

$$X \rightarrow R \equiv (X)_{F_R}^+ = R,$$

- ne obstaja prava podmnožica $X' \subset X$, ki bi funkcionalno določala vse attribute sheme R ,

$$\neg \exists X' (X' \subset X \wedge X' \rightarrow R).$$

Relacijska shema ima lahko več kandidatov za ključ. Pri realizaciji izberemo enega izmed njih, ki ga imenujemo **(primarni) ključ**. Podmnožico atributov relacijske sheme, ki je primarni ključ v neki drugi shemi, imenujemo **tuji ključ**.

Primer. Naj bo $STUDENT = \{Vpisna, Ime, Priimek, Letnik, Posta, Kraj\}$ relacijska shema in $F_{STUDENT}$ pripadajoča množica funkcionalnih odvisnosti. Poiščite vse kandidate za ključ sheme $STUDENT$.

$$F_{STUDENT} = \{Vpisna \rightarrow \{Ime, Priimek, Letnik, Posta\}, Posta \rightarrow Kraj\}$$

$Vpisna$ je edini kandidat za ključ.

Primer. Naj bo $R = ABC$ relacijska shema in F_R pripadajoča množica funkcionalnih odvisnosti. Poiščite vse kandidate za ključ sheme R .

$$F_R = \{A \rightarrow B, B \rightarrow A, AB \rightarrow C, BC \rightarrow A\}$$

Najprej opazimo, da velja $(A)_{F_R}^+ = ABC$. Zato sledi, da je A kandidat za ključ, saj je drugi pogoj iz definicije trivialno izpolnjen. Podobno opazimo, da je tudi B kandidat za ključ.

Iz drugega pogoja definicije sledi, da noben drug kandidat za ključ ne vsebuje A ali B . Možen kandidat je tako zgolj še C , vendar pa velja $(C)_{F_R}^+ = C$.

Kandidata za ključ sta tako le A in B .

8. naloga. Naj bo R relacijska shema, X podmnožica atributov sheme R , $X \subseteq R$, in F_R pripadajoča množica funkcionalnih odvisnosti. Ali je X kandidat za ključ sheme R ?

8.a) $R = ABCDEFG$, $X = EF$, $F_R = \{A \rightarrow B, BE \rightarrow G, D \rightarrow AC, EF \rightarrow A\}$

Ne, ker je $(EF)_{F_R}^+ = ABEFG$.

8.b) $R = ABCDE$, $X = ABE$, $F_R = \{AB \rightarrow C, C \rightarrow A, CD \rightarrow E, E \rightarrow D\}$

Ja, ker je $(ABE)_{F_R}^+ = ABCDE$ in velja minimalnost.

8.c) $R = ABCDEFG$, $X = AG$, $F_R = \{A \rightarrow BCF, B \rightarrow E, E \rightarrow D, G \rightarrow A\}$

Ne, ker ne velja minimalnost čeprav $(AG)_{F_R}^+ = ABCDEFG$.

9. naloga. Naj bo R relacijska shema in F_R pripadajoča množica funkcionalnih odvisnosti. Poiščite vse

kandidate za ključ sheme R .

9.a) $R = AB, F_R = \{A \rightarrow B, B \rightarrow A\}$

| A in B sta kandidata za ključ.

9.b) $R = ABC, F_R = \{\}$

| ABC je edini kandidat za ključ.

9.c) $R = ABC, F_R = \{A \rightarrow B, B \rightarrow A, AB \rightarrow C, BC \rightarrow A, AC \rightarrow B\}$

| A in B sta kandidata za ključ.

9.d) $R = ABCD, F_R = \{A \rightarrow B, B \rightarrow D, D \rightarrow C, C \rightarrow A\}$

| A, B, C in D so kandidati za ključ.

9.e) $R = ABCD, F_R = \{A \rightarrow BCD, AB \rightarrow C, AC \rightarrow B, BC \rightarrow A\}$

| A in BC sta kandidata za ključ.

9.f) $R = ABCDE, F_R = \{A \rightarrow B, B \rightarrow A\}$

| $ACDE$ in $BCDE$ sta kandidata za ključ.

9.g) $R = ABCDE, F_R = \{A \rightarrow BCDE, BCDE \rightarrow A\}$

| A in $BCDE$ sta kandidata za ključ.

9.h) $R = ABCDEFG, F_R = \{A \rightarrow B, B \rightarrow A, ABC \rightarrow DEFG\}$

| AC in BC sta kandidata za ključ.

Formalna normalizacija

Formalna normalizacija je postopek preoblikovanja relacijskih shem oziroma pripadajočih relacij v obliko, pri kateri ne more priti do ažurnih anomalij. Normalizacija uporablja *pristop od spodaj navzgor*. Načrtovanje podatkovne baze tako pričnemo z eno relacijo, ki vsebuje vse pomembne attribute. Relacijo nato postopoma preoblikujemo v želeno normalno obliko:

- **1NO**: prva formalna normalna oblika
- **2NO**: druga formalna normalna oblika
- **3NO**: tretja formalna normalna oblika

- **BCNO**: Boyce-Coddova normalna oblika
- **4(P)NO**: četrta (poslovna) normalna oblika
- **5(P)NO**: peta (poslovna) normalna oblika

1NO, 2NO, 3NO in BCNO temeljijo na **funkcionalnih odvisnostih**, 4NO temelji na večvrednostnih odvisnostih, 5NO pa na stičnih odvisnostih.

Definicija. Relacija je v **prvi normalni obliki** 1NO, če:

- ima določene funkcionalne odvisnosti,
- ima izbran primarni ključ in
- nima večvrednostnih atributov.

Definicija. Relacija je v **drugi normalni obliki** 2NO, če:

- je v prvi normalni obliki 1NO in
- nima parcialnih odvisnosti.

Definicija. Relacija je v **tretji normalni obliki** 3NO, če:

- je v drugi normalni obliki 2NO in
- nima tranzitivnih odvisnosti.

“Each attribute is placed in an entity where it is dependent on the key (1NO), the whole key (2NO) and nothing but the key (3NO), so help me Codd.”

Primer. Naj bo *VPIS* relacijska shema, ki predstavlja vpis študentov v posamezen letnik. Postopoma normalizirajte shemo *VPIS* do tretje normalne oblike 3NO.

VPIS(Vpisna, Ime, Priimek, (ID, ProgramID, ProgramIme, Letnik, Vpisnina))

Shema *VPIS* ni v 1NO.

- 1NO

$\{Vpisna, ID\}$ je edini kandidat za ključ.

VPIS(Vpisna, ID, Ime, Priimek, ProgramID, ProgramIme, Letnik, Vpisnina) z odvisnostmi

$$F_{VPIS} = \{Vpisna \rightarrow \{Ime, Priimek\}, \{Vpisna, ID\} \rightarrow \{ProgramID, Letnik\}, \\ ProgramID \rightarrow ProgramIme, \{ProgramID, Letnik\} \rightarrow Vpisnina\}$$

- 2NO

VPIS(\#Vpisna, ID, ProgramID, ProgramIme, Letnik, Vpisnina) z odvisnostmi

$$F_{VPIS} = \{\{Vpisna, ID\} \rightarrow \{ProgramID, Letnik\}, ProgramID \rightarrow ProgramIme, \\ \{ProgramID, Letnik\} \rightarrow Vpisnina\}$$

$STUDENT(\underline{Vpisna}, Ime, Priimek)$ z odvisnostmi $F_{STUDENT} = \{Vpisna \rightarrow \{Ime, Priimek\}\}$

- 3NO

$VPIS(\#Vpisna, ID, \#ProgramID, \#Letnik)$ z odvisnostmi

$$F_{VPIS} = \{\{Vpisna, ID\} \rightarrow \{ProgramID, Letnik\}\}$$

$STUDENT(\underline{Vpisna}, Ime, Priimek)$ z odvisnostmi

$$F_{STUDENT} = \{Vpisna \rightarrow \{Ime, Priimek\}\}$$

$PROGRAM(\underline{ProgramID}, ProgramIme)$ z odvisnostmi

$$F_{PROGRAM} = \{ProgramID \rightarrow ProgramIme\}$$

$VPISNINA(\#ProgramID, Letnik, Vpisnina)$ z odvisnostmi

$$F_{VPISNINA} = \{\{ProgramID, Letnik\} \rightarrow Vpisnina\}$$

10. naloga. Naj bo $R = ABCD$ relacijska shema, F_R pripadajoča množica funkcionalnih odvisnosti in $X = AB$ primarni ključ R . V kateri normalni obliki je shema R ? Shemo postopoma normalizirajte do 3NO ter za vsako novo shemo navedite ključe in pripadajoče funkcionalne odvisnosti.

$$F_R = \{AB \rightarrow CD, C \rightarrow D\}$$

Shema R je v 2NO.

- 2NO

$R(\underline{A}, \underline{B}, C, D)$ z odvisnostmi $F_R = \{AB \rightarrow CD, C \rightarrow D\}$

- 3NO

$R(\underline{A}, \underline{B}, \#C)$ z odvisnostmi $F_R = \{AB \rightarrow C\}$

$R_1(\underline{C}, D)$ z odvisnostmi $F_{R_1} = \{C \rightarrow D\}$

11. naloga. Naj bo $R = ABCDE$ relacijska shema in F_R pripadajoča množica funkcionalnih odvisnosti. V kateri normalni obliki je shema R ? Shemo postopoma normalizirajte do 3NO ter za vsako novo shemo navedite ključe in pripadajoče funkcionalne odvisnosti.

$$F_R = \{A \rightarrow D, B \rightarrow E, E \rightarrow B\}$$

Shema R ni v 1NO.

- 1NO

ABC in ACE sta kandidata za ključ.

$R(\underline{A}, \underline{B}, \underline{C}, D, E)$ z odvisnostmi $F_R = \{A \rightarrow D, B \rightarrow E, E \rightarrow B\}$

- 2NO

$R(\# \underline{A}, \# \underline{B}, \underline{C})$ z odvisnostmi $F_R = \emptyset$

$R_1(\underline{A}, D)$ z odvisnostmi $F_{R_1} = \{A \rightarrow D\}$

$R_2(\underline{B}, E)$ z odvisnostmi $F_{R_2} = \{B \rightarrow E\}$

- 3NO

Sheme R , R_1 in R_2 so v 3NO.

12. naloga. Naj bo $R = ABCDEF$ relacijska shema, F_R pripadajoča množica funkcionalnih odvisnosti in $X = BC$ primarni ključ R . V kateri normalni obliki je shema R ? Shemo postopoma normalizirajte do 3NO ter za vsako novo shemo navedite ključe in pripadajoče funkcionalne odvisnosti.

$$F_R = \{B \rightarrow DE, C \rightarrow F, E \rightarrow AF\}$$

Shema R je v 1NO.

- 1NO

$R(\underline{A}, \underline{B}, \underline{C}, D, E, F)$ z odvisnostmi $F_R = \{B \rightarrow ADEF, C \rightarrow F, E \rightarrow AF\}$

- 2NO

$R(\# \underline{B}, \underline{C})$ z odvisnostmi $F_R = \emptyset$

$R_1(\underline{A}, \underline{B}, D, E, F)$ z odvisnostmi $F_{R_1} = \{B \rightarrow ADEF, E \rightarrow AF\}$

- 3NO

$R(\# \underline{B}, \underline{C})$ z odvisnostmi $F_R = \emptyset$

$R_1(\underline{B}, D, \# E)$ z odvisnostmi $F_{R_1} = \{B \rightarrow DE\}$

$R_2(\underline{A}, \underline{E}, F)$ z odvisnostmi $F_{R_2} = \{E \rightarrow AF\}$