Uvod
Zavisnosti u relacijskim bazama podataka
Formalni sustavi
Implikacijski problem
Normalne forme
Pitanja?

Teorija baza podataka
Modeliranje i normalizacija baza podataka

Izv. prof. dr. sc. Markus Schatten

Fakultet organizacije i informatike,
Sveučilište u Zagrebu
Pavlinska 2, 42000 Varaždin
markus.schatten@foi.hr

Uvod

Neformalno, bazu podataka možemo karakterizirati na sljedeći način:

Definicija

Baza podataka *je kolekcija podataka, ograničenja i operacija koja reprezentira neke aspekte realnog svijeta.*

Modeliranjem aplikacijske domene AD dolazimo do baze podataka BP. Prema tome, BP je model za AD.

Uvod

- Ostvarenje ili implementacija baze podataka obavlja se primjenom odgovarajućeg sustava za upravljanje bazom podataka.

Definicija

Sustav za upravljanje bazom podataka je poseban programski proizvod (engl. software), a njegova osnovna komponenta je model podataka.

- Prema pripadnom modelu podataka razlikujemo:
  - hijerarhijske sustave za upravljanje bazom podataka (temelje se na hijerarhijskom modelu podataka),
  - mrežne sustave (temelje se na mrežnom modelu podataka),
  - relacijske sustave (temelje se na relacijskom modelu podataka),
  - objektno-orijentirane sustave (temelje se na objektnom modelu podataka),
  - polustrukturirane sustave (temelje se na polustrukturiranom modelu podataka odnosno podatkovnim grafovima) itd.

Uvod

Uvod
Zavisnosti u relacijskim bazama podataka
Formalni sustavi
Implikacijski problem
Normalne forme
Pitanja?

Napredniji sustavi kao što su temporalni sustavi, deduktivni sustavi, objektno/relacijski sustavi, aktivni sustavi, generalizirani (poopćeni) sustavi i sustavi strujanja podataka nastaju proširenjem relacijskog ili drugog modela komponentama za temporalnost, deduktivnost, objektnu orijentiranost, aktivnost, strukturalnost i strujanje.

Model podataka

Definicija

Svaki model podataka, MP, sastoji se od tri komponente, tj., $MP = (S, UI, O)$, gdje je S strukturalna komponenta (kaže u kojem obliku su prikazani podaci), UI je integritetna komponenta (ograničenja na dozvoljena stanja strukture) i O je operativna komponenta (operacije nad strukturama).

Za relacijski model podataka, *RMP*, imamo sljedeće: *S* je skup relacija (tablica), *UI* je skup ograničenja stanja relacija i *O* je skup relacijskih operatora.

Integritetna komponenta

- Zavisnosti u relacijskim bazama podataka predstavljaju važan dio integritetne komponente relacijskog modela.
- Pomoću zavisnosti ograničavamo moguća stanja relacija u bazi podataka.
- Valjanom bazom podataka smatramo onu bazu podataka čije relacije zadovoljavaju propisane uvjete integriteta.
- U oblikovanju sheme relacijske baze podataka posebno važnu ulogu imaju:
  - funkcijske zavisnosti,
  - višeznačne zavisnosti i
  - zavisnosti spoja.

Funkcijske zavisnosti

Slijedi definicija funkcijske zavisnosti.

Definicija

Funkcijska zavisnost

Neka je R relacijska shema, $X, Y \subseteq R$. Izraz $X \rightarrow Y$ je funkcijska zavisnost nad R. $FZ(R) = \{X \rightarrow Y | X, Y \subseteq R\}$ je skup svih funkcijskih zavisnosti nad R. Kažemo da $X \rightarrow Y \in FZ(R)$ vrijedi u relaciji $r(R)$ ako

$$\forall t_1, t_2 \in r \;\; (t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y])$$

Dakle, funkcijska zavisnost $X \rightarrow Y \in FZ(R)$ vrijedi u relaciji $r(R)$ ako je za bilo koja dva sloga t_1 i t_2 iz r ispunjeno da iz jednakosti slogova t_1 i t_2 na skupu atributa X slijedi njihova jednakost i na skupu atributa Y.

Alternativna definicija

Definicija

Neka je R relacijska shema, $X, Y \subseteq R$. Kažemo da $X \to Y \in FZ(R)$ vrijedi u relaciji $r(R)$ ako je pripadno preslikavanje p sa $\Pi_X(r)$ na $\Pi_Y(r)$ funkcija. Ako redove iz $\Pi_X(r)$ nazovemo X-vrijednosti, a redove iz $\Pi_Y(r)$ nazovemo Y-vrijednosti, onda možemo reći da funkcijska zavisnost $X \to Y$ vrijedi u relaciji r ukoliko svakoj X-vrijednosti odgovara točno jedna Y-vrijednost.

Primjer

Primjer

Ispitajmo je li zavisnost $AC \to B$ vrijedi u relaciji

r	A	B	C	D
	1	2	2	3
	2	2	2	2
	1	3	2	3

Višeznačne zavisnosti

Definicija

Višeznačna zavisnost

Neka je $X, Y \subseteq R$. Višeznačna zavisnost nad R je izraz oblika $X \twoheadrightarrow Y$. Kažemo da $X \twoheadrightarrow Y$ vrijedi u $r(R)$ ako

$$\forall t_1, t_2 \in r\{t_1[X] = t_2[X] \Rightarrow \exists t_3 \in r(t_3[XY] = t_1[XY] \wedge t_3[R - XY] = t_2[R - XY])\}$$

Sa $VZ(R)$ označavamo skup svih višeznačnih zavisnosti nad R.

Višeznačne zavisnosti

Iz definicije višeznačne zavisnosti vidimo da $X \twoheadrightarrow Y$ vrijedi u $r(R)$ ako je za bilo koja dva sloga t_1 i t_2 iz r ispunjeno da iz jednakosti slogova t_1 i t_2 na skupu atributa X slijedi postojanje sloga t_3 iz r sa svojstvom da se t_3 podudara sa t_1 na skupu atributa XY i t_3 se podudara sa t_2 na skupu atributa $R - XY$.

Kako su višeznačne zavisnosti posebne zavisnosti spoja, ispitivanje je li višeznačna zavisnost vrijedi u relaciji bit će izvršeno tako da se višeznačna zavisnost pretvori u odgovarajuću zavisnost spoja, a zatim se ispituje dobivena zavisnost spoja u danoj relaciji.

Zavisnosti spoja

Definicija

Zavisnost spoja

Neka su $R_1, \ldots, R_n \subseteq R$ *neprazni skupovi takvi da je* $R_1 \cup \ldots \cup R_n = R$. *Izraz* $\bowtie(R_1, \ldots, R_n)$ *zovemo zavisnost spoja nad* R. *Skup svih zavisnosti spoja nad* R *označavamo sa* $ZS(R)$.
Kažemo da zavisnost spoja $\bowtie(R_1, \ldots, R_n) \in ZS(R)$ *vrijedi u* $r(R)$ *ako je*

$$r = \Pi_{R_1}(r) \bowtie \ldots \bowtie \Pi_{R_n}(r)$$

Primjer

Primjer

Ispitajmo zavisnost spoja $\bowtie(ABC, CD)$ *u relaciji*

r	A	B	C	D
	1	2	2	3
	2	2	2	2
	1	3	2	3

Skupovi zavisnosti

Definicija

Neka je $FVS(R) = FZ(R) \cup VZ(R) \cup ZS(R)$. Za skup zavisnosti $F \subseteq FVS(R)$ kažemo da vrijedi u relaciji $r(R)$ ako svaka zavisnost iz F vrijedi u $r(R)$. Ako F vrijedi u r, onda se r naziva modelom za F.

Relacijska shema

Definicija

Neka je R skup atributa i F skup zavisnosti nad R. Par (R, F) naziva se relacijska shema.

Dakle, relacijska shema (R, F) se sastoji od skupa atributa R i skupa zavisnosti F između atributa iz skupa R.

Uvjet integriteta

Specificirajući (R, F) iskazujemo da će se valjanim (validnim) relacija nad R smatrati samo one relacije nad R koje su modeli za F.

Relacija r se mijenja tijekom vremena kao posljedica ažuriranja. Dozvoljena su samo ona ažuriranja koja čuvaju valjanost relacije r, tj. ažuriranja trebaju biti u skladu sa zadanim uvjetima integriteta F. Slikovito, skup zavisnosti F djeluje kao filter: od svih mogućih stanja relacije r dozvoljena su samo ona stanja koja su modeli za F.

Logička posljedica

Centralni pojam teorije baza podataka je pojam logičke posljedice. Njegova važnost proizlazi iz činjenice da u osnovi gotovo svih temeljnih pojmova (njihovoj karakterizaciji) koristimo pojam logičke posljedice.

Definicija

Logička posljedica

Neka je $F \subseteq FVS(R)$ i $f \in FVS(R)$. f je logička posljedica od F ako za svaku relaciju r nad R vrijedi:

Ako je r model za F, onda je r model i za f.

Činjenicu da je f logička posljedica od F označavamo ovako $F \models f$. Ako f nije logička posljedica od F , onda pišemo $F \not\models f$.

Katalog FZ

U sljedećem katalogu dana su pravila rezoniranja, temeljem logičke posljedice, o funkcijskim zavisnostima.

Katalog FZ

Neka su $X, Y, Z, Y_1 \subseteq R$. Tada vrijedi:

1. $X \rightarrow Y \models XZ \rightarrow YZ$ (proširenje)
2. $X \rightarrow Y, Y \rightarrow Z \models X \rightarrow Z$ (tranzitivnost)
3. $X \rightarrow Y, X \rightarrow Y_1 \models X \rightarrow YY_1$ (unija)
4. $X \rightarrow Y \models X \rightarrow Z$, ako je $Z \rightarrow Y$ (projekcija)
5. $F \models X \rightarrow Y$, gdje je F bilo koji skup funkcijskih zavisnosti i $Y \subseteq X$ (trivijalnost)

Katalog VZ

Pravila rezoniranja o višeznačnim zavisnostima prikazana su u Katalogu VZ.

Katalog VZ

Neka su $X, Z, Z \subseteq R$. Tada vrijedi:

1. $X \twoheadrightarrow Y \models X \twoheadrightarrow (R - XY)$ (komplementiranje)
2. $X \twoheadrightarrow Y \models XW \twoheadrightarrow YV$, gdje je $V \subseteq W \subseteq R$ (proširenje)
3. $X \twoheadrightarrow Y, Y \twoheadrightarrow Z \models X \twoheadrightarrow (Z - Y)$ (tranzitivnost)
4. $X \twoheadrightarrow Y, X \twoheadrightarrow Z \models X \twoheadrightarrow YZ$ (unija)
5. $X \twoheadrightarrow Y, X \twoheadrightarrow Z \models X \twoheadrightarrow (Y \cap Z)$ (p_1)
   $X \twoheadrightarrow Y, X \twoheadrightarrow Z \models X \twoheadrightarrow (Y - Z)$ (p_2)
   $X \twoheadrightarrow Y, X \twoheadrightarrow Z \models X \twoheadrightarrow (Z - Y)$ (p_3)

Pravilo 5. naziva se projekcija za višeznačne zavisnosti i ima tri navedena oblika (p_1), (p_2) i (p_3).

Katalog FVS

Teorija baza podataka Modeliranje i normalizacija baza podataka

Uvod

Zavisnosti u relacijskim bazama podataka

Formalni sustavi

Implikacijski problem

Normalne forme

Pitanja?

Nekoliko pravila rezoniranja, koja povezuju funkcijske zavisnosti, višeznačne zavisnosti i zavisnosti spoja, dana su u Katalogu FVS.

Katalog FVZ

Neka su $X, Y \subseteq R$. Tada vrijedi:

1. $X \rightarrow Y \models X \rightarrow\!\!\!\rightarrow Y$ (obrat ne vrijedi)
2. $X \rightarrow Y \vdash \bowtie(XY, X \cup (R - XY))$
3. $X \rightarrow\!\!\!\rightarrow Y \models \bowtie(XY, X \cup (R - XY))$
4. $\bowtie(XY, X \cup (R - XY)) \models X \rightarrow\!\!\!\rightarrow Y$

Napomene

Točka 1. iz Kataloga FVS kaže da iz funkcijske zavisnosti slijedi višeznačna zavisnost, ali da obrat ne vrijedi. Dalje, točka 2. utvrđuje da funkcijska zavisnosti ima za logičku posljedicu odgovarajuću zavisnost spoja.

Točke 3. i 4. u katalogu FVS kažu da je višeznačna zavisnost $X \twoheadrightarrow Y$ ekvivalentna zavisnosti spoja $\bowtie(XY, X \cup (R - XY))$, jer vrijedi $X \twoheadrightarrow Y \models \bowtie(XY, X \cup (R - XY))$ i $\bowtie(XY, X \cup (R - XY)) \models X \twoheadrightarrow Y$.

Zbog navedenog, ispitivanje višeznačne zavisnosti $X \twoheadrightarrow Y$ u relaciji r, svodi se na ispitivanje korespondentne zavisnosti spoja $\bowtie(XY, X \cup (R - XY))$ u relaciji r.

Ekvivalencija skupa zavisnosti

Definicija

Ekvivalencija skupova zavisnosti

Sa $FVS(R)$ označavamo skup svih funkcijskih, višeznačnih i zavisnosti spoja nad R. Neka su $F, G \subseteq FVS(R)$. Za skup zavisnosti F kažemo da ima za logičku posljedicu skup zavisnosti G ako $F \models g$ za svaku zavisnost g iz G. Tada pišemo $F \models G$.

Ekvivalencija skupova zavisnosti F i G, u oznaci $F \equiv G$, definira se ovako $F \equiv G$ ako $F \models G$ i $G \models F$.

Primjer

Primjer

Ispitajmo da li višeznačna zavisnost $Artikl \rightarrow\rightarrow Boja$ *vrijedi u relaciji*

r	$Artikl$	$Odjel$	$Boja$
	a_1	o_1	$bijela$
	a_1	o_2	$crvena$
	a_1	o_2	$bijela$

Budući je $Artikl \rightarrow\rightarrow Boja \equiv \bowtie(Artikl\ Boja, Artikl\ Odjel)$*, u relaciji* r *treba ispitati zavisnost spoja* $\bowtie(Artikl\ Boja, Artikl\ Odjel)$*.*

Propozicija $\models$

Propozicija

($\models$)

Neka su $F, G, H \subseteq FVS(R)$ skupovi zavisnosti. Tada

1. *$F \models G$, ako je $G \subseteq F$.*
2. *$F \models G$ i $G \models H$ povlači $F \models H$.*

Primjer

Primjer

$R = ABCD$, $F : A \to B, B \to C; G : A \to B, B \to C, A \to C$.
Ispitati $F \equiv G$.

Propozicija $\equiv$

Propozicija

($\equiv$)

Neka su $X, Y, R_1, \ldots, R_k, S \subseteq R$ skupovi atributa. Tada

1. $X \twoheadrightarrow Y \equiv \bowtie(XY, X \cup (R - XY))$
2. $\bowtie(R_1, \ldots, R_k) \equiv \bowtie(R_1, \ldots, R_k, S)$, *gdje je $S \subseteq R_i$ za neki $i = 1, \ldots, k$.*

U točci 1. propozicije ($\equiv$) ponovili smo da je višeznačna zavisnost $X \twoheadrightarrow Y$ ekvivalentna korespondentnoj zavisnosti spoja $\bowtie(XY, X \cup (R - XY))$. Točka 2. ove propozicije kaže da proširenjem zavisnosti spoja $\bowtie(R_1, \ldots, R_k)$ komponentom S koja je podskup neke komponente R_i dobivamo ekvivalentnu zavisnost spoja $\bowtie(R_1, .., R_k, S)$.

Propozicija Isto semantičko ograničenje

Sa $V(R, F)$ označavamo skup svih valjanih relacija nad shemom (R, F).

Propozicija

(Isto semantičko ograničenje)
Neka vrijedi $F \equiv G$. *Tada vrijedi* $V(R, F) = V(R, G)$.

U ovoj propoziciji iskazali smo da ekvivalentni skupovi zavisnosti predstavljaju isto semantičko ograničenje (predstavljaju “isti filter” za valjanost).

Propozicija Kontekstualna neovisnost funkcijskih zavisnosti

Propozicija

(Kontekstualna neovisnost funkcijskih zavisnosti)
Neka je $X, Y \subseteq R$. *Funkcijska zavisnost* $X \rightarrow Y$ *vrijedi u relaciji* $r(R)$ *ako i samo ako* $X \rightarrow Y$ *vrijedi u* $\Pi_{XY}(r)$.

Pojašnjenje

Upravo iskazana propozicija kaže da je za ispitivanje funkcijske zavisnosti $X \to Y$ u relaciji $r(R)$ dovoljno (i nužno) promatrati vrijednosti redova na atributima iz skupa XY; rezultat ne ovisi o vrijednosti redova na preostalim atributima iz $R - XY$.

Za razliku od funkcijske zavisnosti, višeznačna zavisnost je kontekstualno ovisna. Kod ispitivanja višeznačne zavisnosti $X \to\!\!\!\to Y$ u relaciji $r(R)$ nije dovoljno promatrati samo atribute XY, nego je potrebno promatrati i preostale atribute iz $R - XY$. To se može direktno vidjeti iz danog opisa semantike višeznačne zavisnosti, gdje smo rekli da $X \to\!\!\!\to Y$ vrijedi u $r(R)$ ako $\forall t_1, t_2 \in r\{t_1[X] = t_2[X] \Rightarrow \exists t_3 \in r(t_3[XY] = t_1[XY] \wedge t_3[R - XY] = t_2[R - XY])\}$.

Kontekstualna ovisnost višeznačnih zavisnosti opravdava uvođenje ugrađenih višeznačnih zavisnosti.

Ugrađene višeznačne zavisanosti

Ugrađena višeznačna zavisnost definira se kao što slijedi.

Definicija

Ugrađena višeznačna zavisnost
Neka su $X, Y, Z \subseteq R$. Izraz oblika $X \twoheadrightarrow Y/Z$ zovemo ugrađena višeznačna zavisnost nad R. Kažemo da $X \twoheadrightarrow Y/Z$ vrijedi u $r(R)$ ako $X \twoheadrightarrow Y$ vrijedi u $\Pi_{XYZ}(r)$.

Primjer

Uvod
Zavisnosti u relacijskim bazama podataka
Formalni sustavi
Implikacijski problem
Normalne forme
Pitanja?

Primjer

Neka je zadana relacija

r	*Nastavnik*	*Predmet*	*Dijete*	*Datum-rođenja*
	Sedmak	M_1	*Ivan*	$04-04-2001$
	Sedmak	M_1	*Hana*	$01-08-2007$
	Sedmak	M_2	*Ivan*	$04-04-2001$
	Sedmak	M_2	*Hana*	$01-08-2007$
	Tomić	F_1	*Damir*	$05-07-2004$
	Tomić	F_2	*Damir*	$05-07-2004$
	Benić	P_1	*Ivo*	$14-09-2003$

Ispitajmo u relaciji r

(a) *višeznačnu zavisnost* $Nastavnik \twoheadrightarrow Dijete$;

(b) *ugrađenu višeznačnu zavisnost* $Nastavnik \twoheadrightarrow Dijete / Predmet$.

Podskup zavisnosti

- Do sada smo uveli sljedeće zavisnosti: funkcijske zavisnosti, višeznačne zavisnosti, ugrađene višeznačne zavisnosti i zavisnosti spoja.
- Sve one predstavljaju semantičko ograničenje na jednoj relaciji. Zbog toga se kaže da one predstavljaju intrarelacijsko ograničenje.
- U nastavku opisat ćemo jednu novu zavisnost, pod nazivom podskup zavisnost ili inkluzijska zavisnost, koja prestavlja ograničenje na dvije relacije.
- Riječ je dakle o interrelacijskom ograničenju ili međurelacijskom ograničenju. Slijedi uvodni primjer.

Primjer

Primjer

Neka su zadane relacije nastavnik i predaje

nastavnik	$n\#$	$prezime$	$zvanje$
	n_1	Kim	prof
	n_2	Kam	doc
	n_3	Kam	izprof

predaje	$n\#$	$p\#$
	n_1	uz
	n_2	bp2
	n_2	uz

U relaciji nastavnik prikazani su šifra, prezime i zvanje nastavnika, a u relaciji predaje imamo šifre nastavnika i predmeta koje dani nastavnik predaje. Referencijalni integritet nam kaže da u relaciji predaje ne možemo imati nepostojećeg nastavnika tj. nastavnika koji nije naveden u relaciji nastavnik. Formalno, spomenuto ograničenje možemo iskazati zahtjevom da u svim stanjima relacija nastavnik i predaje bude ispunjen uvjet

$$RI : \Pi_{n\#}(predaje) \subseteq \Pi_{n\#}(nastavnik)$$

Izraz RI zove se podskup zavisnost.

Podskup zavisnosti

Definicija

Podskup zavisnosti

Neka su zadani skupovi atributa $X = \{A_1, A_2, \dots, A_k\} \subseteq R_1$ *i* $Y = \{B_1, B_2, \dots, B_k\} \subseteq R_2$. *Neka je dalje* $dom(A_i) = dom(B_i)$ *za svaki* $i \in \{1, 2, \dots, k\}$. *Izraz* $[X, R_1] \subseteq [Y, R_2]$ *zovemo podskup zavisnost.*

Kažemo da podskup zavisnost $[X, R_1] \subseteq [Y, R_2]$ *vrijedi za relacije* $r(R_1)$, $s(R_2)$ *ako je*

$$\Pi_X(r) \subseteq \Pi_Y(s)$$

Podskup zavisnost $[X, R_1] \subseteq [Y, R_2]$ *predstavlja međurelacijsko ograničenje za relacije* r *nad* R_1 *i* s *nad* R_2. *Dozvoljena su samo ona stanja relacije* $r(R_1)$ *i* $s(R_2)$ *u kojima vrijedi podskup zavisnost* $[X, R_1] \subseteq [Y, R_2]$.

Primjer

Primjer

Za relacije nastavnik i predaje iz prethodnog primjera ispitajmo podskup zavisnosti

(a) $[n\#, n\#\ p\#] \subseteq [n\#, n\#\ prezime\ zvanje]$

(b) $[n\#, n\#\ prezime\ zvanje] \subseteq [n\#, n\#\ p\#]$

Zadatak

Zadatak

Ispitajte svaku od zavisnosti iz skupa $F : AB \to D$, $B \twoheadrightarrow AC$, $\bowtie(AB, CD)$, $\bowtie(A, BCD)$ *u relaciji*

r_1	A	B	C	D
	1	2	2	1
	2	1	1	1
	1	2	0	0

Formalni sustavi

Formalni sustav za zavisnosti sastoji se od pravila (zaključivanja) koja omogućuju rezoniranje o zavisnostima.

Definicija

Pravilo zaključivanja (dedukcije) za zavisnosti $FVS(R)$ je izraz oblika

$$f_1, \ldots, f_k \vdash g, \text{ gdje su } f_1, \ldots, f_k, g \in FVS(R)$$

Pravilo $\emptyset \vdash g$ pišemo u obliku $\vdash g$ i nazivamo aksiom.

Formalni sustav za funkcijske zavisnosti

Definicija

Formalni sustav za funkcijske zavisnosti je konačan skup pravila oblika

$$f_1, \ldots, f_k \vdash g, \text{ gdje su } f_1, \ldots, f_k, g \in FZ(R)$$

Definicija

Korektnost formalnog sustava
Formalni sustav $FS : P_1, \ldots, P_m$ *je korektan ako je svako pravilo* P_i *iz* FS *korektno. Pravilo* $P_i : f_1, \ldots, f_k \vdash g$ *je korektno ako vrijedi* $f_1, \ldots, f_k \models g$.

Primjer

Primjer

Ispitajmo korektnost sljedećih pravila:

1. $l_p : X \rightarrow Y \vdash XZ \rightarrow Y$
2. $d_p : X \rightarrow Y \vdash X \rightarrow YZ$

Rješenje #1

1. Pravilo I_p je korektno ako vrijedi $X \to Y \models XZ \to Y$. Rezoniranje, temeljem kataloga FZ, je kao što slijedi:

1. $X \to Y$ zadano
2. $XZ \to YZ$ iz 1. uz proširenje sa Z
3. $YZ \to Y$ trivijalnost
4. $XZ \to Y$ iz 2. i 3. primjenom tranzitivnosti

Prema tome, vrijedi $X \to Y \models XZ \to Y$, tj. pravilo I_p je korektno.

Rješenje #2

2. Pravilo d_p je korektno ako $X \to Y \models X \to YZ$.

Neka je $R = ABC$, $X = A$, $Y = B$, $Z = C$ i

r	A	B	C
	1	2	3
	1	2	2

Lako provjerimo da $A \to B$ vrijedi u r, a $A \to BC$ ne vrijedi u r. Prema tome, $A \to B \not\models A \to BC$, tj. $X \to Y \not\models X \to YZ$. Dobiveni rezultat nam kaže da pravilo d_p nije korektno.

Derivabilnost zavisnosti

Notacija $[FS]\ f_1, \ldots, f_k \vdash g$ indicira da je zavisnost g derivabilna iz zavisnosti $f_1, \ldots, f_k$ primjenom formalnog sustava FS. Navedeno znači da postoji niz zavisnosti $g_1, \ldots, g_m$ sa svojstvima:

1. $g_m = g$
2. Svaki g_i je jednak nekom članu u nizu $f_1, \ldots, f_k$ ili je dobiven iz prethodnih g-ova u nizu $g_1, \ldots, g_m$ primjenom nekog pravila iz FS.

Potpunost formalnog sustava

Definicija

(Potpunost formalnog sustava) Formalni sustav FS je potpun za funkcijske zavisnosti ako

$$(\forall F \subseteq FZ(R))(\forall g \subseteq FZ(R)[F \models g \Rightarrow [FS]\ F \vdash g]$$

Prema tome, formalni sustav FS je potpun za funkcijske zavisnosti ako za svaki skup funkcijskih zavisnosti F i za svaku funkcijsku zavisnost g vrijedi: ako je g logička posljedica od F, onda se g može deducirati iz F primjenom pravila formalnog sustava FS.

Armstrongov formalni sustav (AFS)

Definicija

AFS je formalni sustav za funkcijske zavisnosti, a sastoji se od sljedećih pravila:

A1: $\vdash X \to Y$, *ako je* $Y \subseteq X$ *(trivijalnost)*

A2: $X \to Y \vdash XZ \to YZ$ *(proširenje)*

A3: $X \to Y, Y \to Z \vdash X \to Z$ *(tranzitivnost)*

Propozicija Korektnost i potpunost AFS

Propozicija

(Korektnost i potpunost AFS) Armstrongov formalni sustav AFS je korektan i potpun za funkcijske zavisnosti.

Zatvarač skupa atributa

Definicija

Neka je $F \subseteq FZ(R)$ i $X \subseteq R$. Zatvarač skupa atributa X s obzirom na skup funkcijskih zavisnosti F, u oznaci X_F^+, definiramo ovako:

$$X_F^+ = \{A \in R | [AFS] F \vdash X \to A\}$$

Propozicija Dedukcija pomoću zatvarača

Propozicija

(Dedukcija pomoću zatvarača) $[AFS]F \vdash X \to Y$ *ako i samo ako* $Y \subseteq X_F^+$.

Z-primjenjivost

Algoritam za računanje X_F^+ koristi pojam zatvarač- primjenjivosti ili z-primjenjivosti.

Definicija

Zavisnost $X_1 \rightarrow Y_1$ je z-primjenjiva na X ako je $X_1 \subseteq X$ i $Y_1 \nsubseteq X$.

Algoritam X_F^+

Algoritam

(Računanje X_F^+)
Ulaz: $F \subseteq FZ(R)$, $X \subseteq R$
Izlaz: X_F^+
Postupak:

1. *Staviti $X_0 = X$*
2. *Odrediti zavisnost iz F koja je z-primjenjiva na X_0.*
   Ako takva zavisnost ne postoji, onda $X_F^+ = X_0$.
   Ako je $V \to W$ izabrana zavisnost iz F, koja je z-promjenjiva na X_0, onda izračunati $X_1 = X_0 \cup W$
3. *Primijeniti korak 2. na X_1.*
   Navedeni postupak treba ponavljati sve dotle dok ne dobijemo takav skup atributa X_i za koji ne postoji niti jedna zavisnost u F koja je z-primjenjiva na X_i. Dobiveni X_i je X_F^+.

Primjer

Primjer

$R = ABCD,\ F : A \to B,\ BC \to D,\ D \to A$

Izračunajmo (a) A_F^+ *(b)* B_F^+

Propozicija Potpunost AFS

Propozicija

*(**potpunost AFS**)*
Formalni sustav AFS je potpun za funkcijske zavisnosti.

Formalni sustav za funkcijske i višeznačne zavisnosti

Definicija

Formalni sustav za funkcijske i višeznačne zavisnosti (FS), sastoji se od sljedećih pravila:

1. *A1*
2. *A2*
3. *A3 (iz definicije AFS)*
4. $X \twoheadrightarrow Y \vdash X \twoheadrightarrow (R - XY)$
5. $X \twoheadrightarrow Y \vdash XW \twoheadrightarrow YV$, *gdje je* $V \subseteq W$
6. $X \twoheadrightarrow Y, Y \twoheadrightarrow Z \vdash X \twoheadrightarrow (Z - Y)$
7. $X \rightarrow Y \vdash X \twoheadrightarrow Y$
8. $X \twoheadrightarrow Y, W \rightarrow Z \vdash X \rightarrow Z$, *gdje je* $Z \subseteq Y$ *i* $W \cap Y = \emptyset$.

Propozicija Korektnost i potpunost FS

Propozicija

FS je korektan i potpun za funkcijske i višeznačne zavisnosti.

Ponovimo da korektnost znači mogućnost zamjene svakog pojavljivanja simbola $\vdash$ sa simbolom $\models$, a potpunost znači mogućnost zamjene simbola $\models$ sa simbolom $\vdash$.

Implikacijski problem

Centralni problem teorije baza podataka je implikacijski problem (I-problem):

Ispitati da li vrijedi $F \models f$, gdje je $F \subseteq FVS(R)$ i $f \in FVS(R)$.

Postupci rješavanja I-problema

(Se) Semantički (koristi se definicija logičke posljedice $\models$ ili katalozi FZ, VZ i FVS)

(Si) Sintaksni (primjena formalnih sustava)

(Al) Algoritamski (primjenjuju se algoritmi koji koriste X_F^+, Bazu zavisnosti, Chase-postupak).

U primjeni teorije baza podataka veliku važnost ima treći postupak Al na koji ćemo se fokusirati.

Rješavanje I-problema za funkcijske zavisnosti

Rješavanje I-problema za funkcijske zavisnosti dano je sljedećim algoritmom.

Algoritam

Algoritam testira $F \models f$*, gdje je* $F \subseteq FZ(R)$ *i* $f \in FZ(R)$
Ulaz: $F \subseteq FZ(R)$, $f : X \to Y \in FZ(R)$
Izlaz:

DA *ako* $F \models f$

NE *ako* $F \not\models f$

Postupak:

1. *Izračunati* X_F^+.
2. *Ako* $Y \subseteq X_F^+$ *, onda* $F \models f$*. Ako nije* $Y \subseteq X_F^+$*, onda* $F \not\models f$

Primjer

Primjer

$R = ABCD$, $F:\ A \to B, A \to CD, B \to A$

Riješimo I-problem $F \models B \to AD$.

Rješavanje I-problem za funkcijske i višeznačne zavisnosti

U rješavanju I-problema za funkcijske i višeznačne zavisnosti koristimo algoritam koji se temelji na konceptu baze zavisnosti.

Definicija

Baza zavisnosti

Neka je $F \subseteq FVZ(R)$, $X \subseteq R$. *Neka je* V_z *skup višeznačnih zavisnosti definiran ovako:*

1. *Svaka višeznačna zavisnost iz* F *je u* V_z.
2. *Za svaku funkcijsku zavisnost* $X \to A_1 \dots A_k$ *iz* F, *u* V_z *staviti višeznačne zavisnosti* $X \twoheadrightarrow A_1, \dots, X \twoheadrightarrow A_k$.

Baza zavisnosti od X *s obzirom na* V_z, $BZ(X, V_z)$, *je particija od* $R - X$, *dobivena algoritmom (Baza zavisnosti).*

Algoritam - Baza zavisnosti

Algoritam

(Baza zavisnosti)
Ulaz: $X \subseteq R$, $V_z \subseteq VZ(R)$.
Izlaz: $BZ(X, V_z)$
Postupak:

1. $X_0 = R - X$
2. *Izabrati zavisnost* $V \twoheadrightarrow W \in V_z$ *tako da bude* $V \cap X_0 = \varnothing$ *i* $W \cap X_0 \neq \varnothing$. *Ako takva zavisnost ne postoji, onda je* $BZ(X, V_z) = X_0$, *inače*
3. $X_1 = X_0 \cap W$, $X_2 = X_0 - W$.
4. *Primijeniti 2. na* X_1 *i* X_2.
5. *Navedeni postupak ponavljati dok ne dođemo do skupova atributa* $T_1, \ldots, T_m$ *koje više nije moguće dekomponirati.*
6. $BZ(X, V_z) = \{T_1, \ldots, T_m\}$

Algoritam (B)

Algoritma koji rješava I-problem za funkcijske i višeznačne zavisnosti.

Algoritam

(B)
Algoritam testira $F \models f$, *gdje je* $F \subseteq FVZ(R)$, $f \in FVZ(R)$.
Ulaz: $F \subseteq FVZ(R)$, $f : X \to Y$ *ili* $f : X \twoheadrightarrow Y$, *gdje su* $X, Y \subseteq R$, $X \cap Y = \varnothing$.
Izlaz:

$$\begin{array}{ll} \text{DA} & \text{ako } F \models f \\ \text{NE} & \text{ako } F \not\models f \end{array}$$

Postupak:

1. *Transformirati* F *u* V_z.
2. *Izračunati* $BZ(X, V_z)$.
3. $F \models X \twoheadrightarrow Y$ *ako i samo ako* Y *je unija nekih članova iz* $BZ(X, V_z)$.

   $F \models X \to A$ *ako i samo ako*

   1. $A \in BZ(X, V_z)$ *i*
   2. *Postoji* $X_i \to Y_i$ *u* F *takva da* $A \cap X_i = \varnothing$ *i* $A \in Y_i$.

Napomena

Kako vrijedi $X \to A_1 \dots A_k \equiv X \to A_1, \dots, X \to A_k$, rješavajući I-problem $F \models X \to A$, možemo riješiti i I-problem $F \models X \to Y$, gdje je $Y = A_1 A_2 \dots A_k$.

Primjer

Uvod
Zavisnosti u relacijskim bazama podataka
Formalni sustavi
Implikacijski problem
Normalne forme
Pitanja?

Primjer

$R = ABCD$ $F: A \rightarrow B, BC \rightarrow D, B \twoheadrightarrow C$

Riješimo I-probleme

1. $F \models A \rightarrow C$
2. $F \models A \twoheadrightarrow C$

Rješavanje I-problem za funkcijske, višeznačne i zavisnosti spoja

Teorija baza podataka
Modeliranje i normalizacija baza podataka

Uvod

Zavisnosti u relacijskim bazama podataka

Formalni sustavi

Implikacijski problem

Normalne forme

Pitanja?

Prelazimo sada na rješavanje I-problema u općem slučaju, gdje su uključene funkcijske zavisnosti, višeznačne zavisnosti i zavisnosti spoja.

Algoritam se temelji na modifikacijama (transformacijama) tablica kojima se nastoji ‘uloviti’ ciljni red odnosno ciljni stupac.

Navedeni postupak se naziva Chase-postupak.

Početna tablica T_0

Definicija

Početna tablica T_0

Za relacijsku shemu (R, F), *gdje je* $R = A_1 A_2 \dots A_m$, *neka je zadana dekompozicija od R, u oznaci* $d(R)$, $d(R) = R_1, R_2, \dots, R_n$. *Pretpostavka da je* $d(R)$ *dekompozicija od R znači da vrijedi*

(a) $R_i \neq \emptyset$, *za svako* $i = 1, 2, \dots, n$

(b) (b) $R_1 \cup R_2 \cup \dots \cup R_n = R$.

Tablica T_0 *definira se ovako:*

$$\begin{array}{c|ccc} T_0 & A_1 & \dots & A_m \\ \hline R_1 & & & \\ \vdots & & T_{ij} & \\ R_n & & & \end{array} \qquad T_{ij} = \begin{cases} a_j & \text{ako } A_j \in R_i \\ b_{ij}, & \text{ako } A_j \notin R_i \end{cases}$$

Pri čemu je $i = 1, \dots, n$ *i* $j = 1, \dots, m$.

Ciljni redak i ciljni stupac

Prema tome, tablica T_0 sastoji se od odgovarajućih simbola a_j i b_{ij}. Simbole a_j zovemo istaknuti simboli. Redovi tablice T_0 imenovani su skupovima (komponentama) iz dekompozicije $d(R)$. Vidi se da T_0 ima n redova i m stupaca. Red koji se sastoji samo od istaknutih simbola zovemo ciljni red, tj. red

$$cr : a_1 a_2 \dots a_m \text{ je ciljni red}$$

Dalje, stupac koji se sastoji samo od istaknutih simbola zove se ciljni stupac (oznaka je cs). Uočite da ciljni stupac cs za atribut A_j ima oblik

$$\begin{array}{c|c} cs & A_j \\ \hline & a_j \\ & a_j \\ & \vdots \\ & a_j \end{array}$$

Dakle, ciljni stupac cs za atribut A_j sastoji se od pojavljivanja istog simbola a_j.

Modifikacije

Definicija

FZ, VZ, ZS modifikacije

***(FZ)** Modifikacija temeljem funkcijske zavisnosti*

Neka je $f : X \to Y$;

Zapis $T_{i+1} = Mo[T_i, f]$ *znači da je tablica* T_{i+1} *dobivena iz* T_i *transformacijom temeljenoj na funkcijskoj zavisnosti* f. *Transformacija se sprovodi ovako:*

Za bilo koja dva reda $t_1, t_2 \in T_i$ *takva da je* $t_1[X] = t_2[X]$ *izjednačite* $t_1[Y]$ *sa* $t_2[Y]$ *koristeći pravilo (FZ) koje glasi ovako:*

- *Ako su oba simbola oblika* a_j, *nema potrebe za promjenom;*
- *Ako je jedan simbol* a_j, *a drugi* b_{ij}, *tada* b_{ij} *zamijeniti sa* a_j;
- *Ako je jedan simbol oblika* b_{ij}, *a drugi* b_{kj}, *pri čemu je* $k < i$ *tada zamijeniti* b_{ij} *sa* b_{kj}.

Primjer

Primjer

Neka je $\boxed{\begin{array}{llll} t_1[Y]: & a_3 & b_{57} & b_{59} \\ t_2[Y]: & b_{13} & b_{17} & a_9 \end{array}}$

Novi $t_1[Y]$ *i* $t_2[Y]$ *poprimaju oblik* $\boxed{\begin{array}{llll} t_1[Y]: & a_3 & b_{17} & a_9 \\ t_2[Y]: & a_3 & b_{17} & a_9 \end{array}}$

Dakle, pravilo (FZ) ne mijenja simbol a_j, *a simbol* b_{ij} *se zamjenjuje sa* a_j *ili sa* b_{kj}, *gdje je* $k < i$.

Modifikcije

Definicija

***(ZS)** Modifikacija temeljem zavisnosti spoja*

Neka je $f : \bowtie(R_1, R_2, \dots, R_m)$. Zapis $T_{i+1} = Mo[T_i, f]$ znači da je tablica T_{i+1} dobivena iz T_i transformacijom temeljenoj na zavisnosti spoja f.

Transformaciju sprovodimo koristeći pravilo (ZS):

$$T_{i+1} = \Pi_{R_1}(T_i) \bowtie \Pi_{R_2}(T_i) \bowtie \dots \bowtie \Pi_{R_m}(T_i)$$

Modifikacije

Definicija

(VS) *Modifikacija temeljem višeznačne zavisnosti*
Neka je $f : X \twoheadrightarrow Y$. *Kako je* $f \equiv \bowtie(XY, X(R - XY))$, *transformaciju temeljem* f *sprovodimo primjenom zavisnosti spoja* $\bowtie(XY, X(R - XY))$.

Napomena

Neka je $F \subseteq FVS(R)$ skup zavisnosti. Ako primjenimo sve moguće modifikacije koristeći zavisnosti iz F, krećući od početne tablice T_0, u oznaci $Mo[T_0, F]$, dobivamo konačan niz tablica $T_0, T_1, \ldots, T_f$. Finalna tablica T_f ima svojstvo da ju više ne mijenja niti jedna modifikacija po zavisnostima iz F. Uočite da u dobivenoj tablici T_f vrijede sve zavisnosti iz F. Prema tome, T_f je model za F.

Algoritam C

Algoritam

(C)

Algoritam testira $F \models f$, *gdje je* $F \subseteq FVS(R)$, $f \in FVS(R)$.
Ulaz: $F \subseteq FVS(R)$, $f \in FVS(R)$ *Izlaz:*

DA *ako* $F \vDash f$

NE *ako* $F \nvDash f$

Postupak:
Postupak ima tri modula (C_1, C_2 *i* C_3) *Modul* C_1 *rješava I-problem* $F \models \bowtie(R_1, \ldots, R_m)$, C_2 *rješava I-problem* $F \models X \rightarrow\!\!\rightarrow Y$, *te* C_3 *rješava I-problem* $F \models X \rightarrow Y$.

Algoritam C

Algoritam

(Cont.)

C_1

Testiramo $F \models \bowtie(R_1, \dots, R_m)$ *(Zavisnosti spoja)*

Postupak:

1. *Napisati* T_0 *za zadani* R *i* $d(R) = R_1, \dots, R_m$.
2. *Izračunati* T_f.
3. $F \models \bowtie(R_1, \dots, R_m)$ *ako i samo ako* $cr \in T_f$.

Algoritam C

Algoritam

(Cont.)

$\mathbf{C_2}$

Testiramo $F \models X \twoheadrightarrow Y$ *(Višeznačne zavisnosti)*

Postupak:

1. *Napisati* T_0 *za zadani* R *i* $d(R) = XY, X \cup (R - XY)$
2. *Izračunati* T_f
3. $F \models X \twoheadrightarrow Y$ *ako i samo ako* $cr \in Tf$.

Algoritam C

Algoritam

(Cont.)

C_3

Testiramo $F \models X \to Y$ *(Funkcijske zavisnosti)*

Postupak:

1. *Napisati* T_0 *za zadani* R *i* $d(R) = X, R$.
2. *Računati* T_f.
3. $F \models X \to Y$ *ako i samo ako su svi* Y*-stupci ciljni, tj. sastoje se samo od odgovarajućih simbola* a_j.

Primjer

Primjer

Neka je $R = ABCD$, $F : A \to C, C \twoheadrightarrow B, \bowtie(AB, C, BD)$.
Riješimo I-probleme:

(a) $F \models \bowtie(ABC, BD)$
(b) $F \models B \twoheadrightarrow A$
(c) $F \models A \to B$

Zadatak

Zadatak

Neka je $R = ABCD$, $F : A \rightarrow C, C \rightarrow\!\!\!\rightarrow B, \bowtie(AB, C, BD)$.

Riješite I-probleme:

(a) $F \models A \rightarrow B$

(b) $F \models B \rightarrow\!\!\!\rightarrow A$

(c) $F \models \bowtie(ABC, BD)$.

Normalne forme

Logičko oblikovanje sheme relacijske baze podataka temelji se na postupku normalizacije relacijske sheme (R, F), koji rezultira shemom relacijske baze podataka u odgovarajućoj normalnoj formi. U postupku normalizacije primjenjuju se određena svojstva zavisnosti i svojstva dekompozicije relacijske sheme.

Svojstva zavisnosti

Definicija

Trivijalna i netrivijalna zavisnost

Neka je zadana relacijska shema (R, F). *Za funkcijsku zavisnost* $X \to Y \in FVS(R)$ *kažemo da je trivijalna ako je* $Y \subseteq X$. *U protivnome, ako* $Y \not\subseteq X$, *onda kažemo da je* $X \to Y$ *netrivijalna zavisnost.*

Svojstva zavisnosti

Definicija

Trivijalna i netrivijalna zavisnost (Cont.)

Višeznačna zavisnost $X \twoheadrightarrow Y \in FVS(R)$ *je trivijalna ako je* $Y \subseteq X$ *ili* $XY = R$*. U protivnome, ako nije (*$Y \subseteq X$ *ili* $XY = R$*), onda kažemo da je* $X \twoheadrightarrow Y$ *netrivijalna višeznačna zavisnost. Uočite da netrivijalnost višeznačne zavisnosti* $X \twoheadrightarrow Y$ *znači da vrijedi* $Y \not\subseteq X$ *i* $XY \neq R$*.*

Svojstva zavisnosti

Definicija

Trivijalna i netrivijalna zavisnost (Cont.)

Neka su $R_1, \ldots, R_n \subseteq R$ neprazni skupovi takvi da je $R_1 \cup \ldots \cup R_n = R$. Skup $d(R) = \{R_1, \ldots, R_n\}$ nazvali smo dekompozicijom od R, a članove dekompozicije zovemo komponente dekompozicije. Uz konvenciju o kraćem zapisu skupa, dekompoziciju $d(R)$ možemo zapisati ovako $d(R) = R_1, \ldots, R_n$.
Izraz $\bowtie(R_1, \ldots, R_n)$ nazvali smo zavisnost spoja nad R.
Zavisnost spoja $\bowtie(R_1, \ldots, R_n) \in FVS(R)$ je trivijalna ako je $R_i = R$ za neko $i = 1, \ldots, n$. U protivnome, ako je $R_i \neq R$ za svako $i = 1, \ldots, n$, onda kažemo da je zavisnost spoja $\bowtie(R_1, \ldots, R_n)$ netrivijalna.

Primjer

Primjer

Neka je zadana relacijska shema (R, F), *gdje je* $R = ABCDE$, $F : AB \to BD, ABC \to BC, C \twoheadrightarrow AD, B \twoheadrightarrow ACDE, \bowtie(AB, CD, DE), \bowtie(AB, ABCDE, DE)$.

Vrijede sljedeća svojstva zavisnosti iz F:

1. $AB \to BD$ *je netrivijalna, jer* $BD \not\subseteq AB$
2. $ABC \to BC$ *je trivijalna, jer* $BC \not\subseteq ABC$
3. $C \twoheadrightarrow AD$ *je netrivijalna, jer* $AD \not\subseteq C$ *i* $C \cup AD \neq R$
4. $B \twoheadrightarrow ACDE$ *je trivijalna, jer* $B \cup ACDE = R$
5. $\bowtie(AB, CD, DE)$ *je netrivijalna, jer* $AB \neq R$, $CD \neq R$, $DE \neq R$
6. $\bowtie(AB, ABCDE, DE)$ *je trivijalna, jer je druga komponenta,* $ABCDE$, *jednaka* R.

Propozicija Trivijalnost

Propozicija

Trivijalnost

Neka je $f \in FVS(R)$ trivijalna zavisnost. Tada f vrijedi u bilo kojoj relaciji $r(R)$.

Iz navedene propozicije proizlazi da trivijalna zavisnost ne predstavlja nikakvo semantičko ograničenje. U dizajniranju sheme (R, F) nema smisla u F uključivati trivijalne zavisnosti.

Parcijalna zavisnost

Definicija

Parcijalna zavisnost

Neka je zadana je shema (R, F), $F \subseteq FVS(R)$. *Funkcijska zavisnost* $X \to Y \in FVS(R)$ *je parcijalna s obzirom na F ako* $\exists Z \subset X : F \models Z \to Y$.

Tranzitivna funkcijska zavisnost

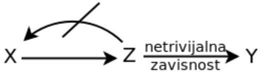
Definicija

Tranzitivna funkcijska zavisnost

Neka je zadana je shema (R, F), $F \subseteq FVS(R)$. *Funkcijska zavisnost* $X \to Y \in FVS(R)$ *je tranzitivna s obzirom na* F *ako* $\exists Z \subseteq R : F \models X \to Z, F \models Z \to Y$, *gdje je* $Z \to Y$ *netrivijalna zavisnost,* $F \not\models Z \to X$.

Tranzitivni dijagram

Činjenica da je $X \to Y$ tranzitivna zavisnost znači mogućnost konstrukcije sljedećeg tranzitivnog dijagrama:



Ključ

Koncept ključa relacijske sheme igra važnu ulogu u postupku normalizacije. Definira se kao što slijedi.

Definicija

Ključ

Neka je zadana relacijska shema (R, F). *Za skup atributa* $K \subseteq R$ *kažemo da je ključ za relacijsku shemu* (R, F) *ako vrijedi:*

(k_1) $F \models K \to R$ *(jedinstvena identifikacija)*

(k_2) *Ne postoji* $K_1 \subset K : F \models K_1 \to R$ *(minimalnost)*

Ključni i neključni atributi

Definicija

Atribut $A \in R$ je ključni atribut ako je član nekog ključa sheme (R, F); atribut $A \in R$ je neključni atribut ako nije član niti jednog ključa sheme (R, F).

Propozicija Ključni trik

Propozicija

Ključni trik

Neka je (R, F) relacijska shema, gdje je F skup funkcijskih zavisnosti nad R. Neka je dalje $X_k \subseteq R$ skup onih atributa iz R koji nisu u desnoj strani niti jedne funkcijske zavisnosti iz F. Tada svaki ključ K za (R, F) sadrži X_k, tj. $X_k \subseteq K$ za bilo koji K. Ako je X_k ključ za (R, F), onda je to jedini ključ.

Primjer

Primjer

Zadana je relacijska shema (R, F), *gdje je* $R = ABCDE$, $F : AB \to D$, $C \to AD$, $B \to D$.

(a) *Odredite sve ključeve za* (R, F);

(b) *Je li zavisnost* $AB \to D$ *parcijalna s obzirom na* F?

(c) *Je li zavisnost* $C \to D$ *tranzitivna s obzirom na* F?

(d) *Postoji li parcijalna zavisnost neključnog atributa od ključa?*

(e) *Postoji li tranzitivna zavisnost neključnog atributa od ključa?*

Određivanje ključa

Algoritam

(ključ)
Ulaz: (R, F)
Izlaz: Ključ K *za* (R, F)
Postupak

1. *Odrediti skup atributa* $X_{nd} = \{A \in R | A$ *nije na desnoj strani niti jedne zavisnosti iz* $F\}$;
   1. *Ako je* $X_{nd} \neq \varnothing$, *onda izračunajte* $(X_{nd})^+_F$. *Ukoliko je* $(X_{nd})^+_F = R$, *tada je* X_{nd} *jedini ključ za* (R, F). *Ako je* $(X_{nd})^+_F \neq R$, *onda* X_{nd} *nije ključ; daljnjim svim mogućim proširivanjem* X_{nd}, *uz čuvanje minimalnosti, dobiju se svi ključevi za* (R, F) *(koji sadrže* X_{nd}*).*
   2. *Ako je* $X_{nd} = \varnothing$, *onda ključni trik nije primjenjiv, te se prelazi na korak 2.*
2. *Izabrati skup atributa* $Y \subseteq R$ *i izračunati* Y^+_F.
   1. *Ako je* $Y^+_F = R$, *onda* Y *ima svojstvo* (k_1) *ključa; izbacivanjem atributa iz* Y, *uz kontrolu minimalnosti, dolazimo do skupa* Y_k, $Y_k \subseteq Y$, *iz kojeg ne možemo dalje izbacivati atribute bez narušavanja svojstva* (k_1). *Dobiveni skup* Y_k *je ključ za* (R, F). *Svi mogući različiti oblici izbacivanja atributa iz skupa* Y *rezultiraju svim ključevima koji su sadržani u* Y.
   2. *Ako je* $Y^+_F \neq R$, *onda skup* Y *nema svojstvo* (k_1)*; svim mogućim proširivanjem skupa* Y, *uz čuvanje minimalnosti, dobiju se svi ključevi za* (R, F) *koji sadrže skup* Y.
3. *Svim mogućim izborima skupa* Y *u točki 2., nalaze se svi ključevi za* (R, F).

Primjer

Primjer

Primijenimo algoritam (ključ) u određivanju svih ključeve za relacijsku shemu (R, F)*, gdje je*

$$R = ABCDF : A \to B, BC \to A, A \to CD$$

Svojstva dekompozicije relacijske sheme

Podsjetimo se da je $d(R) = R_1, \ldots, R_k$ dekompozicija relacijske sheme (R, F) ako su ispunjeni uvjeti

(a) $R_i \neq \emptyset$ za svako $i = 1, 2, \ldots, k$

(b) $R_1 \cup \ldots \cup R_k = R$

Za kvalitetan logički dizajn sheme relacijske baze podataka poželjno je da dekompozicija ima svojstvo čuvanja informacije i čuvanja zavisnosti.

Čuvanje informacije

Definicija

Čuvanje informacije
Za dekompoziciju $d(R_1, \ldots, R_k)$ relacijske sheme (R, F) kažemo da čuva informaciju ako vrijedi $F \models \bowtie(R_1, \ldots, R_k)$.

Navedena definicija kaže da dekompozicija d čuva informaciju ako za svaku relaciju r nad R vrijedi: ako je r model za F, onda je r model i za zavisnost spoja $\bowtie(R_1, \ldots, R_k)$, tj. r se dobije prirodnim spojem svojih projekcija na komponente dekompozicije. Kažemo da se r može restaurirati iz svojih projekcija na komponente dekompozicije.

Primjer

Primjer

$R = ABC, F : A \to B, C \to B, d(R) : CB, CA.$ *Ispitati da li d čuva informaciju.*

Potrebno je riješiti implikacijski problem $F \models \bowtie(CB, CA)$.

Primjer

Primjer

$R = ABC$, $F : A \to B, C \to B, d(R) : CB, CA$. *Ispitati da li d čuva informaciju.*

Potrebno je riješiti implikacijski problem $F \models \bowtie(CB, CA)$.

Čuvanje zavisnosti

Definicija

Čuvanje zavisnosti

Za dekompoziciju $d(R): R_1, \ldots, R_k$ relacijske sheme (R, F) kažemo da čuva zavisnosti F ako za svaku relaciju r nad R vrijedi: ako je $\Pi_{R_1}(r)$ model za $\Pi_{R_1}(F)$ i ... $\Pi_{R_k}(r)$ model za $\Pi_{R_k}(F)$, onda je r model za F.

U definiciji čuvanja zavisnosti imamo skupove zavisnosti $\Pi_{R_i}(F)$, $i = 1, \ldots, k$, koji se nazivaju projekcije skupa zavisnosti F na komponente dekompozicije $R_1, \ldots, R_k$. Dakle, dekompozicija $d(R)$ čuva zavisnosti ako je za svaku relaciju $r(R)$ ispunjeno: iz činjenice da $\Pi_{R_i}(F)$ vrijedi u $\Pi_{R_i}(r)$ za svako $i = 1, \ldots, k$ slijedi da F vrijedi u r.
Navedeno možemo iskazati nešto kraće i na ovaj način: iz valjanosti projekcija $\Pi_{R_i}(r)$ (lokalna valjanost) proizlazi valjanost $r(R)$ (globalna valjanost).

Projekcija skupa zavisnosti

Definicija

Projekcija skupa zavisnosti

Projekcija skupa funkcijskih zavisnosti F

$$\Pi_{R_i}(F) = \{X \rightarrow Y \mid F \models X \rightarrow Y \text{ } i \text{ } XY \subseteq Ri\}$$

Projekcija skupa višeznačnih zavisnosti

$$\Pi_{R_i}(F) = \{X \twoheadrightarrow Y \mid \exists Z \subseteq R : F \models X \twoheadrightarrow Z, Y = Z \cap R_i), XY \subseteq Ri\}$$

Primjer

Primjer

$R = ABCDE$, $R_1 = ACD$, $F : A \to BC$, $B \to DE$, $B \twoheadrightarrow C$, $AC \twoheadrightarrow DE$

Ispitati koje od zavisnosti: $A \to BC$, $A \to DE$, $AC \twoheadrightarrow DE$, $AC \twoheadrightarrow D$, $C \twoheadrightarrow D$ *pripadaju projekciji* $\Pi_{R_1}(F)$.

Primjer

Rješenje Ispitujemo $A \to BC$ Kako vrijedi $ABC \not\subseteq R_1$, zaključujemo da $A \to BC \notin \Pi_{R_1}(F)$.
Ispitujemo $A \to D$ Imamo $A \to BC, B \to DE \models A \to D$. Zato, $F \models A \to D$. Kako je i $AD \subseteq R_1$, zaključujemo da vrijedi $A \to D \in \Pi_{R1}(F)$.
Ispitujemo $AC \twoheadrightarrow DE$ Imamo $ACDE \not\subseteq R_1$. Zato, $AC \twoheadrightarrow DE \notin \Pi_{R_1}(F)$.
Ispitujemo $AC \twoheadrightarrow D$ Iz činjenica $ACD \subseteq R_1$ i $F \models AC \twoheadrightarrow D$ proizlazi $AC \twoheadrightarrow D \in \Pi_{R_1}(F)$.
Ispitujemo $C \twoheadrightarrow AD$ Imamo $CAD \subseteq R_1$. Dalje, vrijedi $F \models C \twoheadrightarrow ABDE$, jer je $C \twoheadrightarrow ABDE$ trivijalna višeznačna zavisnost za $R = ABCDE$. Budući je $ABDE \cap R_1 = AD$, zaključujemo da vrijedi $C \twoheadrightarrow AD \in \Pi_{R_1}(F)$. Iz posljednjeg rezultata vidimo da iako $F \not\models C \twoheadrightarrow AD$, ipak vrijedi da je $C \twoheadrightarrow AD$ u projekciji $\Pi_{R_1}(F)$.

Primjer

Uvod
Zavisnosti u relacijskim bazama podataka
Formalni sustavi
Implikacijski problem
Normalne forme
Pitanja?

Rješenje Ispitujemo $A \to BC$ Kako vrijedi $ABC \not\subseteq R_1$, zaključujemo da $A \to BC \notin \Pi_{R_1}(F)$.
Ispitujemo $A \to D$ Imamo $A \to BC, B \to DE \models A \to D$. Zato, $F \models A \to D$.
Kako je i $AD \subseteq R_1$, zaključujemo da vrijedi $A \to D \in \Pi_{R1}(F)$.
Ispitujemo $AC \twoheadrightarrow DE$ Imamo $ACDE \not\subseteq R_1$. Zato, $AC \twoheadrightarrow DE \notin \Pi_{R_1}(F)$.
Ispitujemo $AC \twoheadrightarrow D$ Iz činjenica $ACD \subseteq R_1$ i $F \models AC \twoheadrightarrow D$ proizlazi $AC \twoheadrightarrow D \in \Pi_{R_1}(F)$.
Ispitujemo $C \twoheadrightarrow AD$ Imamo $CAD \subseteq R_1$. Dalje, vrijedi $F \models C \twoheadrightarrow ABDE$, jer je $C \twoheadrightarrow ABDE$ trivijalna višeznačna zavisnost za $R = ABCDE$. Budući je $ABDE \cap R_1 = AD$, zaključujemo da vrijedi $C \twoheadrightarrow AD \in \Pi_{R_1}(F)$. Iz posljednjeg rezultata vidimo da iako $F \not\models C \twoheadrightarrow AD$, ipak vrijedi da je $C \twoheadrightarrow AD$ u projekciji $\Pi_{R_1}(F)$.

Primjer

Rješenje Ispitujemo $A \to BC$ Kako vrijedi $ABC \not\subseteq R_1$, zaključujemo da $A \to BC \notin \Pi_{R_1}(F)$.
Ispitujemo $A \to D$ Imamo $A \to BC, B \to DE \models A \to D$. Zato, $F \models A \to D$. Kako je i $AD \subseteq R_1$, zaključujemo da vrijedi $A \to D \in \Pi_{R1}(F)$.
Ispitujemo $AC \twoheadrightarrow DE$ Imamo $ACDE \not\subseteq R_1$. Zato, $AC \twoheadrightarrow DE \notin \Pi_{R_1}(F)$.
Ispitujemo $AC \twoheadrightarrow D$ Iz činjenica $ACD \subseteq R_1$ i $F \models AC \twoheadrightarrow D$ proizlazi $AC \twoheadrightarrow D \in \Pi_{R_1}(F)$.
Ispitujemo $C \twoheadrightarrow AD$ Imamo $CAD \subseteq R_1$. Dalje, vrijedi $F \models C \twoheadrightarrow ABDE$, jer je $C \twoheadrightarrow ABDE$ trivijalna višeznačna zavisnost za $R = ABCDE$. Budući je $ABDE \cap R_1 = AD$, zaključujemo da vrijedi $C \twoheadrightarrow AD \in \Pi_{R_1}(F)$. Iz posljednjeg rezultata vidimo da iako $F \not\models C \twoheadrightarrow AD$, ipak vrijedi da je $C \twoheadrightarrow AD$ u projekciji $\Pi_{R_1}(F)$.

Primjer

Rješenje Ispitujemo $A \to BC$ Kako vrijedi $ABC \not\subseteq R_1$, zaključujemo da $A \to BC \notin \Pi_{R_1}(F)$.
Ispitujemo $A \to D$ Imamo $A \to BC, B \to DE \models A \to D$. Zato, $F \models A \to D$.
Kako je i $AD \subseteq R_1$, zaključujemo da vrijedi $A \to D \in \Pi_{R1}(F)$.
Ispitujemo $AC \twoheadrightarrow DE$ Imamo $ACDE \not\subseteq R_1$. Zato, $AC \twoheadrightarrow DE \notin \Pi_{R_1}(F)$.
Ispitujemo $AC \twoheadrightarrow D$ Iz činjenica $ACD \subseteq R_1$ i $F \models AC \twoheadrightarrow D$ proizlazi $AC \twoheadrightarrow D \in \Pi_{R_1}(F)$.
Ispitujemo $C \twoheadrightarrow AD$ Imamo $CAD \subseteq R_1$. Dalje, vrijedi $F \models C \twoheadrightarrow ABDE$, jer je $C \twoheadrightarrow ABDE$ trivijalna višeznačna zavisnost za $R = ABCDE$. Budući je $ABDE \cap R_1 = AD$, zaključujemo da vrijedi $C \twoheadrightarrow AD \in \Pi_{R_1}(F)$. Iz posljednjeg rezultata vidimo da iako $F \not\models C \twoheadrightarrow AD$, ipak vrijedi da je $C \twoheadrightarrow AD$ u projekciji $\Pi_{R_1}(F)$.

Primjer

Rješenje Ispitujemo $A \rightarrow BC$ Kako vrijedi $ABC \not\subseteq R_1$, zaključujemo da $A \rightarrow BC \notin \Pi_{R_1}(F)$.
Ispitujemo $A \rightarrow D$ Imamo $A \rightarrow BC, B \rightarrow DE \models A \rightarrow D$. Zato, $F \models A \rightarrow D$.
Kako je i $AD \subseteq R_1$, zaključujemo da vrijedi $A \rightarrow D \in \Pi_{R1}(F)$.
Ispitujemo $AC \rightarrow\!\!\!\rightarrow DE$ Imamo $ACDE \not\subseteq R_1$. Zato, $AC \rightarrow\!\!\!\rightarrow DE \notin \Pi_{R_1}(F)$.
Ispitujemo $AC \rightarrow\!\!\!\rightarrow D$ Iz činjenica $ACD \subseteq R_1$ i $F \models AC \rightarrow\!\!\!\rightarrow D$ proizlazi $AC \rightarrow\!\!\!\rightarrow D \in \Pi_{R_1}(F)$.
Ispitujemo $C \rightarrow\!\!\!\rightarrow AD$ Imamo $CAD \subseteq R_1$. Dalje, vrijedi $F \models C \rightarrow\!\!\!\rightarrow ABDE$, jer je $C \rightarrow\!\!\!\rightarrow ABDE$ trivijalna višeznačna zavisnost za $R = ABCDE$. Budući je $ABDE \cap R_1 = AD$, zaključujemo da vrijedi $C \rightarrow\!\!\!\rightarrow AD \in \Pi_{R_1}(F)$. Iz posljednjeg rezultata vidimo da iako $F \not\models C \rightarrow\!\!\!\rightarrow AD$, ipak vrijedi da je $C \rightarrow\!\!\!\rightarrow AD$ u projekciji $\Pi_{R_1}(F)$.

Čuvanje funkcijskih zavisnosti

Propozicija

Čuvanje funkcijskih zavisnosti

Neka je zadana shema (R, F), gdje je $F \subseteq FZ(R)$. Dekompozicija $d(R) : R_1, \dots, R_k$ čuva zavisnosti F ako i samo ako

$$\Pi_{R_1}(F) \cup \dots \cup \Pi_{R_k}(F) \models F$$

Algoritam za testiranje čuvanja funkcijskih zavisnosti

Algoritam koristi koncept R_i-operacije na skupu atributa X s obzirom na skup funkcijskih zavisnosti F.

Definicija

R$_i$ *operacija*

R_i-operacija na skupu atributa X definirana je kao:

$$R_i(X, F) = X \cup [(X \cap R_i)^+_F \cap R_i]$$

Primjer

Primjer

$R = ABCD$, $X = AC$, $R_1 = AB$, $F : A \to B,\ B \to C$. Izračunati $R_1(X, F)$.

Rješenje

$$
\begin{aligned}
AB(AC, F) &= AC \cup [(AC \cap AB)_F^+ \cap AB] \\
&= AC \cup [(A)_F^+ \cap AB] \\
&= AC \cup [ABC \cap AB] \\
&= AC \cup AB \\
&= ABC
\end{aligned}
$$

Algoritam za provjeru čuvanja funkcijskih zavisnosti

Algoritam

(Čuvanje funkcijskih zavisnosti)
Testiramo $G \models F$*, gdje je* $G = \Pi_{R_1}(F) \cup \ldots \cup \Pi_{R_k}(F)$*. G se neće eksplicite računati!*
Ulaz: (R, F), $F \subseteq FZ(R)$, $d(R) : R_1, \ldots, R_k$
Izlaz:

- ***Da** ako d čuva F*
- ***Ne** ako d ne čuva F*

Postupak:

1. *Izabrati zavisnost* $f : X \rightarrow Y$ *iz F te ispitati da li d čuva f; Ispitivanje koristi postupak:*
   1. $X_0 = X$
   2. $X_{i+1} = R_j(X_i, F)$ *za neki* $R_j \in d(R)$
   3. *Prvi* X_k *za koji vrijedi* $R_t(X_k, F) = X_k$*, za svaki* $R_t \in d(R)$*, je* X_G^+
   4. $G \models X \rightarrow Y$*, tj. d čuva* $X \rightarrow Y$ *ako i samo ako* $Y \subseteq X_G^+$.
2. *d čuva F ako i samo ako d čuva svaku zavisnost iz F.*

Primjer

Primjer

$R = ABCD$, $d(R) : AB, BC, CD$; $F : A \to B$, $B \to C$, $C \to D$, $D \to A$. Ispitati je li dekompozicija d čuva F.

Rješenje I

Imamo $G = \Pi_{AB}(F) \cup \Pi_{BC}(F) \cup \Pi_{CD}(F)$. Budući je $A \to B \in \Pi_{AB}(F)$, slijedi $G \models A \to B$, tj. zavisnost $A \to B$ je sačuvana. Analogno možemo zaključiti da su sačuvane i zavisnosti $B \to C$ i $C \to D$. Za ispitivanja čuvanja zavisnosti $D \to A$ trebamo računati R_i - operacije.

$$\boxed{\begin{aligned} CD(D,F) &= D \cup [(D \cap CD)^+_F \cap CD] \\ &= D \cup [D^+_F \cap CD] \\ &= D \cup [DABC \cap CD] \\ &= D \cup CD \\ &= CD \end{aligned}}$$

Kako u dobivenom rezultatu nemamo sadržanu desnu stranu zavisnosti $D \to A$, računamo sljedeću R_i - operaciju na dobivenom rezultatu.

Rješenje II

$$\boxed{\begin{aligned} BC(CD,F) &= CD \cup [(CD \cap BC)^+_F \cap BC] \\ &= CD \cup [(C)^+_F \cap BC] \\ &= CD \cup [CDAB \cap BC] \\ &= CDB \end{aligned}}$$

Sljedeća R_i - operacija daje

$$\boxed{\begin{aligned} AB(CDB,F) &= CDB \cup [(CDB \cap AB)^+_F \cap AB] \\ &= CDB \cup [(B)^+_F \cap AB] \\ &= CDB \cup [BCDA \cap AB] \\ &= CDB \cup AB \\ &= ABCD \end{aligned}}$$

Rješenje III

Prema tome, $D_G^+ = ABCD$. Budući je $A \subseteq D_G^+$, možemo zaključiti da vrijedi $G \models D \to A$, što i znači da je zavisnost $D \to A$ sačuvana dekompozicijom $d(R)$. Budući su sve zavisnosti iz F sačuvane, zaključujemo da dekompozicija $d(R)$ čuva zavisnosti F.

Slabosti relacijske sheme

U sljedećem primjeru analiziramo slabosti relacijske sheme i ukazujemo na postupak otklanjanja uočenih slabosti.

Primjer

Neka je zadana relacijska shema (R, F), gdje je $R =$ Artikl#, Dobavljač#, Grad; F : Dobavljač# $\rightarrow$ Grad. Neka relacija do (dobavljač) nad R u nekom trenutku vremena ima sljedeći oblik

do	Artikl#	Dobavljač#	Grad
	a_1	d_1	Zagreb
	a_2	d_1	Zagreb
	a_3	d_1	Zagreb
	a_2	d_3	Sisak
	a_4	d_2	Rijeka

*Sadržaj relacije do dan je semantikom: do(a, d, g) znači da dobavljač d lociran u gradu g dobavlja artikl a.
Prema tome, prvi red u tablici do ima sljedeću interpretaciju: dobavljač d_1 lociran u gradu Zagreb dobavlja artikl a_1.
Relacija do nema 'dobru' strukturu (shemu) za predstavljanje navedenih podataka, jer sadrži redundantno ponavljanje grada dobavljača za svaki artikl.*

Anomalije

Definicija

Anomalije

(AB) Anomalija brisanja *Brisanjem redova za artikle a1, a2, a3 (kada d1 prestane dobavljati artikle), gubimo informaciju o dobavljaču i pripadnom gradu (nepoželjan efekt).*

(AU) Anomalija upisivanja *Nemogućnost upisivanja dobavljača i pripadnog grada ako dani dobavljač trenutno ne dobavlja neki artikl.*

(AM) Anomalija modifikacije *Nemogućnost nezavisnog ažuriranja vrijednosti atributa Grad.*

Probleme (AB) i (AU) ne možemo riješiti pomoću NULL znaka tj. upisujući odgovarajuće parcijalne redove.

Uzrok problema

Uzrok problema je što postoji parcijalna zavisnost neključnog atributa Grad o ključu $K=$ Artikl#, Dobavljač#. Navedeno znači da (R,F) nije u $2NF$.

Problem redundancije, a samim tim i anomalija ažuriranja, možemo riješiti eliminacijom parcijalne zavisnosti neključnog atributa Grad o ključu $K=$ Artikl#, Dobavljač# tako da izvršimo $2NF$ dekompoziciju relacijske sheme (R,F). Dekomponirajući R, koristeći funkcijsku zavisnost Dobavljač# $\to$ Grad, dobivamo:

$$d(R): R_1 = \text{Dobavljač\#, Grad} \quad R_2 = \text{Dobavljač\#, Atrikl\#}.$$

Ovdje koristimo činjenicu
Dobavljač# $\to$ Grad $\models \bowtie$(Dobavljač Grad, Dobavljač Artikl).

Dekompozicija

Općenito, pripadnu dekompoziciju preko funkcijske zavisnosti $X \to Y$ možemo prikazati dijagramom

$$fz: \qquad X \to Y$$

$$d(R): XY \qquad\qquad X(R - XY)$$

Shema relacijske baze podataka poprima oblik:

$$SRBP: (R_1, \Pi_{R_1}(F)), \quad (R_2, \Pi_{R_2}(F))$$

$SRBP$ je u $2NF$, tj., svaka relacijska shema iz $SRBP$ je u $2NF$.

Rješenje

Umjesto relacije *do*, baza podataka *BP*, nad relacijskom shemom *SRBP*, sastoji se od projekcija *do* na R_1 i R_2. Tako dobivamo $BP : do_1, do_2$, gdje je $do_1 = \Pi_{\text{Dobavljač\# Grad}}(do)$ i $do_2 = \Pi_{\text{Dobavljač\# Artikl\#}}(do)$. Relacije do_1 i do_2 izgledaju ovako:

do_1	Dobavljač#	Grad
	d_1	Zagreb
	d_3	Sisak
	d_2	Rijeka

do_2	Dobavljač#	Artikl#
	d_1	a_1
	d_1	a_2
	d_1	a_3
	d_3	a_2
	d_2	a_4

Uočite da u dobivenoj bazi podataka $BP = do_1, do_2$ nemamo prije opisane anomalije ažuriranja. Također, primijetite da ponavljanje dobavljača d_1 za svaki od artikala koje d_1 dobavlja (u relaciji do_2) nije redundantno, jer inače ne bismo znali što nam sve dobavlja d_1.

Svojstva dekompozicije

Dekompozicija $d(R) : R_1, R_2$ ima sljedeća svojstva:

1. $d(R)$ je $2NF$ dekompozicija
2. $d(R)$ čuva informaciju ($do = do_1 \bowtie do_2$)
3. $d(R)$ čuva skup zavisnosti F , tj., $\Pi_{R_1}(F) \cup \Pi_{R_2}(F) \models F$.

Normalne forme

- Normalne forme su određena ograničenja koja treba zadovoljavati relacijska shema u cilju njene kvalitete.
- Opisat ćemo sljedeće normalne forme: 1NF, 2NF, 3NF, BCNF, 4NF, 5NF i 6NF.
- Svaka viša normalna forma predstavlja strožije ograničenje i rezultira u izvjesnom smislu, koji će biti opisan kasnije, kvalitetniji dizajn sheme relacijske baze podataka.
- Postupak transformacije relacijske sheme (R, F) u odgovarajuću shemu relacijske baze podataka, zasniva se na dekomponiranju (R, F), naziva se normalizacija, koja čini osnovu logičkog oblikovanja sheme relacijske baze podataka.

Logičko oblikovanje baze podataka

Definicija

Problem logičkog oblikovanja sheme relacijske baze podataka

Za danu relacijsku shemu (R, F) treba odrediti dekompoziciju $d(R) : R_1, \dots, R_m$, tj., shemu relacijske baze podataka $SRBP : (R_1, \Pi_{R_1}(F)), \dots, (R_m, \Pi_{R_m}(F))$, tako da bude ispunjeno:

1. *SRBP je u željenoj normalnoj formi,*
2. *$d(R)$ čuva informaciju,*
3. *$d(R)$ čuva zavisnosti F.*

Jednostavni i složeni objekti

Definicija

Jednostavni i složeni objekti

Neka je $A = \{a, b, c, \ldots\}$ skup jednostavnih objekata. Nazivamo ih još elementarnim ili atomaranim objektima. Koristeći konstrukte za skupove, liste, grafove i relacije, dobivamo skupove složenih ili neelemenatarnih objekata (neatomarni objekti).

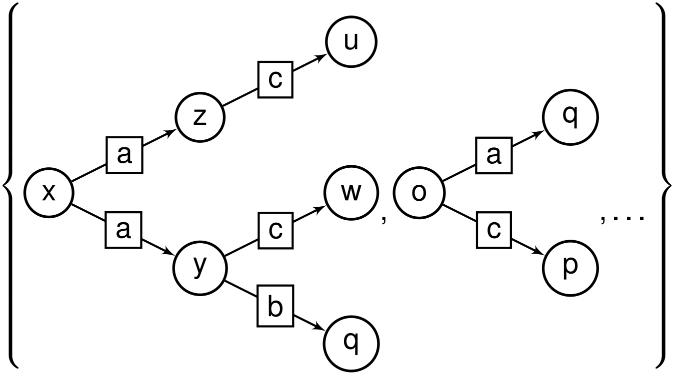
Primjer I

Uvod
Zavisnosti u relacijskim bazama podataka
Formalni sustavi
Implikacijski problem
Normalne forme
Pitanja?

$$
\begin{aligned}
A &= \mathit{Alf} \cup I \cup R \cup \mathrm{Dat} \cup \mathrm{Novac};\\
S &= \{\{a, b\}, \emptyset, \ldots\};\\
L &= \{\langle a, b, a, c\rangle, \langle\rangle, \ldots\};\\
R &= \left\{ \begin{array}{c|cc} r_1 & A & B \\ \hline & 1 & 2 \\ & 3 & 1 \end{array}, \begin{array}{c|ccc} r_2 & B & C & D \\ \hline & 4 & 1 & 1 \\ & 7 & 5 & 1 \end{array}, \ldots \right\};
\end{aligned}
$$

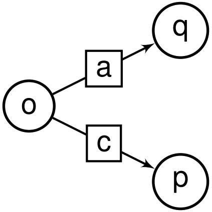
Primjer II

$G =$



Primjer III

$$K = \left\{ \left\langle \begin{array}{c|cc} r_1 & A & B \\ \hline & 1 & 2 \\ & 3 & 1 \end{array}, \{a, b\}, \text{(graf)} \right\rangle, \ldots \right\}$$



A je skup atomskih objekata (sadrži alfanumerički tip, cjelobrojni tip, realne brojeve, datum i td.). *S*, *L*, *R*, i *K* se sastoje od složenih objekata: *S* je skup čiji elementi su skupovi, *L* je skup čiji elementi su liste, *R* je skup koji se sastoji od relacija, *G* je skup koji se sastoji od grafova, a *K* je skup koji se sastoji od elemenata koji su izgrađeni kombinacijom elemenata iz skupova *S*, *L*, *R* i *G*.

Atomski i složeni atributi

Definicija

Atomski i složeni atributi

Atribut A je atomski (jednostavan) atribut ako se $Dom(A)$ sastoji samo od atomskih objekata; u protivnome, A je složen atribut.

1NF i PNF

Definicija

1NF

(R, F) *je u* $1NF$ *ako su svi atributi iz* R *atomski atributi.*

Definicija

PNF

(R, F) *je u poopćenoj* NF *ako je svaki atribut iz* R *jednostavan ili složen.*

U daljem pretpostavljamo da je (R, F) je u $1NF$, gdje je $F \subseteq FVS(R)$.

2NF

Definicija

2NF

(R, F) je u $2NF$ ako s obzirom na F ne postoji parcijalna zavisnost neključnog atributa od ključa.

3NF

Definicija

3NF

(R, F) je u $3NF$ ako s obzirom na F ne postoji tranzitivna zavisnost neključnog atributa od ključa.

Zatvarač skupa zavisnosti

U nastavku karakteriziramo pojam zatvarača skupa zavisnosti, nužnog za definiciju ostalih normalnih formi.

Definicija

***Zatvarač skupa zavisnosti* F**

Zatvarač skupa zavisnosti F, u oznaci F^+, dan je jednakošću

$$F^+ = \{f \in FVS(R) | F \models f\}$$

Skup zavisnosti F^+ predstavlja sve one zavisnosti koje su logička posljedica od F.

BCNF

Definicija

BCNF

(R, F) je u BCNF ako za svaku netrivijalnu funkcijsku zavisnost $X \to Y \in F^+$ vrijedi: lijeva strana X sadrži ključ od (R, F).

4NF

Definicija

4NF

(R, F) je u $4NF$ ako za svaku netrivijalnu višeznačnu zavisnost $X \twoheadrightarrow Y \in F^+$ vrijedi: lijeva strana X sadrži ključ od (R, F).

5NF

Definicija

5NF

(R, F) je u 5NF ako za svaku netrivijalnu zavisnost spoja $\bowtie(R_1, \ldots, R_k) \in F^+$ vrijedi: svaka komponenta R_i, $i = 1, \ldots, k$, sadrži neki ključ od (R, F).

6NF

Definicija

6NF

(R, F) *je u* $6NF$ *ako u* F^+ *nema netrivijalnih zavisnosti spoja.*

Propozicija Odnos normalnih formi

Propozicija

Odnos normalnih formi

Svaka viša normalna forma povlači nižu normalnu formu.

Formalno, Ako je (R, F) u nNF, onda je (R, F) u mNF, gdje je $m < n$.

(Pretpostavljamo uređaj: $1 < 2 < 3 < BCNF < 4 < 5 < 6$)

3NF dekompozicija

Propozicija

3NF dekompozicija

Za svaku shemu (R, F), gdje je $F \subseteq FZ(R)$, postoji $3NF$ dekompozicija koja čuva informaciju i zavisnosti.

Dokaz je dan algoritmom sinteze 3NF preko kanonskog pokrivača.

Algoritam sinteze 3NF

Algoritam sinteze 3NF omogućuje sintezu sheme relacijske baze podataka, koja je u 3NF, čuva informaciju i zavisnosti. Algoritam koristi kanonski pokrivač skupa funkcijskih zavisnosti.

Definicija

***Kanonski pokrivač od* F**

Kanonski pokrivač od F, $kp(F)$, dobije se iz F primjenom sljedeća tri koraka u navedenom redoslijedu:

1. *Desna dekompozicija*
2. *Lijeva redukcija*
3. *Izbacivanje redundantnih (suvišnih) zavisnosti*

Slikovito, navedenu proceduru transformacije skupa zavisnosti F u kanonski pokrivač kp(F) možemo prikazati sljedećim dijagramom

$$F \xrightarrow{[1] \quad [2] \quad [3]} kp(F)$$

Desna dekompozicija

Definicija

[1] Desna dekompozicija

Desno dekomponiranje se temelji na ekvivalenciji

$$X \to A_1A_2 \dots A_m \equiv X \to A_1, X \to A_2, \dots, X \to A_m$$

Kažemo da smo desnim dekomponiranjem zavisnosti $X \to A_1A_2 \dots A_m$ *dobili skup zavisnosti* $X \to A_1, X \to A_2, \dots, X \to A_m$*, gdje se desne strane sastoje od jednog atributa, koji je ekvivalentan polaznoj zavisnosti* $X \to A_1A_2 \dots A_m$*. Ako sa dd označimo operator desne dekompozicije, onda navedenu transformaciju možemo zapisati ovako* $dd(X \to A_1A_2 \dots A_m) = X \to A_1, X \to A_2, \dots, X \to A_m$*. Neka je* $F = \{f_1, f_2, \dots, f_n\}$ *skup funkcijskih zavisnosti. Desna dekompozicija od* F *je* $dd(F) = \{dd(f_1), dd(f_2), \dots, dd(f_n)\}$*.*

Propozicija **dd**

Propozicija

dd

$$dd(F) \equiv F$$

Prema tome, desnom dekompozicijom skupa funkcijskih zavisnosti F dobivamo skup zavisnosti $dd(F)$ koji je ekvivalentan sa F. Pri tome, sve zavisnosti iz $dd(F)$ imaju na svojoj desnoj strani jedan atribut.

Primjer

Uvod
Zavisnosti u relacijskim bazama podataka
Formalni sustavi
Implikacijski problem
Normalne forme
Pitanja?

Primjer

Neka je $F = A \to B, B \to CD, CD \to ABC$. *Tada je* $dd(F) = A \to B, B \to C, B \to D, CD \to A, CD \to B, CD \to C$.

Lijeva redukcija

Definicija

[2] Lijeva redukcija

Lijeva redukcija funkcijske zavisnosti $X \to Y$ s obzirom na skup zavisnosti F, u oznaci $lr(F, X \to Y)$, je zavisnost $X_1 \to Y$ takva da vrijedi

(a) $X_1 \subseteq X$

(b) $F \models X_1 \to Y$

(c) *ne postoji* $X_2 \subset X_1 : F \models X_2 \to Y$

Pišemo $lr(F, X \to Y) = X_1 \to Y$ ili kraće $lr(X \to Y) = X_1 \to Y$ kada se F podrazumijeva.
Uvjet (b) kaže da je zavisnost $X_1 \to Y$ lijevo reducirana odnosno da nema suvišnih atributa u lijevoj strani X_1.
Neka je $F = \{f_1, f_2, \ldots, f_n\}$ skup funkcijskih zavisnosti. Lijeva redukcija od F je

$$lr(F) = \{lr(f_1), lr(f_2), \ldots, lr(f_n)\}$$

Propozicija **Ir**

Propozicija

Ir

$Ir(F) \equiv F$

Nestandardna zavisnost

Definicija

Nestandardna zavisnost

Zavisnosti oblika $\varnothing \to Y$ zovu se nestandardne funkcijske zavisnosti.

Propozicija

nestandardna zavisnost

Neka F ne sadrži nestandardne zavisnosti. Tada je svaka funkcijska zavisnost, čija lijeva strana sadrži samo jedan atribut, lijevo reducirana s obzirom na F.

Primjer

Teorija baza podataka Modeliranje i normalizacija baza podataka

Uvod

Zavisnosti u relacijskim bazama podataka

Formalni sustavi

Implikacijski problem

Normalne forme

Pitanja?

Primjer

$F : A \to B, B \to CD, AC \to B, A \to D$.

Odredimo $lr(F)$.

Izbacivanje suvišnih zavisnosti

Definicija

[3] Izbacivanje suvišnih zavisnosti

Za zavisnost $f \in F$ kažemo da je suvišna (redundantna) u F ako f slijedi iz preostalih zavisnosti u F, tj. ako vrijedi $F - \{f\} \models f$. Za skup zavisnosti F kažemo da je redundantan skup ako ima suvišnih zavisnosti; u protivnome, F je neredundantan skup. U koraku [3] izbacuju se sve suvišne zavisnosti; rezultat je neredundantan skup zavisnost.

Notacija $irz(F)$ označuje skup zavisnosti sa sljedećim svojstvima:

(a) $irz(F) \subseteq F$

(b) $irz(F)$ *je neredundantan skup zavisnosti*

Dakle, $irz(F)$ je rezultat primjene postupka [3] na skup zavisnosti F.

Primjer

Primjer

Neka je $F : A \to B, B \to C, B \to D, A \to B, A \to D$. *Odredimo* $irz(F)$. *Izbacivanje suvišnih zavisnosti provest ćemo krećući se po* F *u smjeru od prve prema zadnjoj zavisnosti.*

Propozicija **irz**

Propozicija

irz

$$irz(F) \equiv F$$

Propozicija (irz) kaže da izbacivanje suvišnih zavisnosti čuva ekvivalenciju tj. rezultirajući skup $irz(F)$ je ekvivalentan polaznom skupu F. Nakon kompletiranja opisa postupaka [1], [2] i [3], zaključimo da svaki od spomenutih postupaka čuva ekvivalenciju, a to onda ima za posljedicu da vrijedi $kp(F) \equiv F$. Dalje, $kp(F)$ je pojednostavljena reprezentacija skupa F. Naziv kanonski pokrivač ukazuje da riječ o osnovnom ili standardnom obliku skupa zavisnosti temeljem kojeg ćemo, kao što će se poslije vidjeti, dobiti relativno jednostavan i praktično vrlo važan algoritam sinteze 3NF.

Primjer

Primjer

$F: A \to BC, B \to C, AB \to D$. Odredimo $kp(F)$.

Sinteza 3NF

Algoritam

Sinteza 3NF

Ulaz: $(R, F), F \subseteq FZ(R)$

Izlaz: dekompozicija $d(R) : R_1, \ldots, R_k$ *sa svojstvima:*

1. $d(R)$ *je 3NF dekompozicija, tj.* $(R_1, P[R_1](F)), \ldots, (R_k, P[R_k](F))$ *su u 3NF;*
2. $d(R)$ *čuva informaciju;*
3. $d(R)$ *čuva zavisnosti.*

Postupak:

1. *Izračunati* $kp(F)$*;*
2. *Sintetizirati komponente* $R_1, \ldots, R_k$*;*
3. *Eventualno dodati ključ za* (R, F) *kao novu komponentu dekompozicije;*
4. *Smanjiti broj komponenti dekompozicije.*

Točka 1. postupka je već objašnjena. Preostaje opisati preostale točke 2., 3. i 4.

Sinteza komponenti

Definicija

2. sinteza komponenti temeljem *kp*(F)

Neka je $kp(F)$: $\boxed{\begin{array}{ccc} X_1 & \to & A_1 \\ X_2 & \to & A_2 \\ & \vdots & \\ X_k & \to & A_k \end{array}}$ *Tada je* $\boxed{\begin{array}{ccc} R_1 & = & X_1A_1 \\ R_2 & = & X_2A_2 \\ & \vdots & \\ R_k & = & X_kA_k \end{array}}$

Vidimo da se komponente sastoje od atributa korespondentnih zavisnosti iz $kp(F)$.

Eventualno dodavanje ključa

Definicija

3. eventualno dodavanje ključa za $(\mathbf{R}, \mathbf{F})$

Ukoliko niti jedna od komponenti $R_1, R_2, \ldots, R_k$ iz prethodne točke ne sadrži ključ za (R, F), onda se računa jedan od ključeva za (R, F) i dodaje kao nova komponenta. Time se dobiva dekompozicija $d(R) : R_1, R_2, \ldots, R_k, K$, gdje je K ključ za (R, F).

Smanjenje broja komponenti dekompozicije

Definicija

4. smanjenje broja komponenti dekompozicije

Smanjenje komponenti dekompozicije možemo postići na dva načina:

1. *u koraku 2. umjesto da svakoj zavisnosti iz* $kp(F)$ *pridružimo jednu komponentu, možemo svakoj grupi zavisnosti iz* $kp(F)$ *koje imaju istu lijevu stranu pridružiti jednu komponentu, koja se sastoji od zajedničke lijeve strane i preostalih atributa na desnim stranama zavisnosti iz grupe. Tako, na primjer, ako je* $kp(F) : X \to A, X \to B, Y \to C, Y \to D, Y \to E$*, onda je* $R_1 = XAB, R_2 = YCDE$.
2. *eliminiramo iz dekompozicije sve one komponente koje su podskupovi neke druge komponente iz dekompozicije. Na primjer, ako je* $d(R) : AB, ABCD, CDE, CDEF$*, onda ćemo nakon eliminacije podskupova dobiti* $d_1(R) : ABCD, CDEF$.

Primjer

Primjer

Neka je zadana relacijska shema (R, F)*, gdje je* $R = ABCDE$,
$F : AB \to C, C \to B, CD \to A$
Primijenimo algoritam sinteze 3NF.

Propozicija nestandardna zavisnost i normalne forme

Propozicija

nestandardna zavisnost i normalne forme

Neka je (R, F) relacijska shema i neka F ne sadrži nestandardne zavisnosti (zavisnosti oblika $\emptyset \to Y$). Tada vrijedi

1. *Svaka dvokomponentna relacijska shema $(A_1 A_2, P[A_1, A_2](F))$ je u 2NF, 3NF, BCNF;*
2. *(R, F) je u 2NF ako su svi ključevi od (R, F) jednočlani (sastoje se samo od jednog atributa).*

Primjer

Primjer

(BCNF)
Neka je zadana relacijska shema (R, F), *gdje je* $R = ABC$,
$F : AB \rightarrow C, C \rightarrow B$.

(a) *ispitati je li* (R, F) *u* $BCNF$

(b) *ako* (R, F) *nije u* $BCNF$, *onda odredite* $BCNF$ *dekompoziciju koja čuva informaciju*

Rješenje I

Rješenje

(a) Temeljem propozicije (ključni trik) dobivamo $A \subseteq K$, gdje je K bilo koji ključ za (R, F). Vrijedi $F \models C \to B$. Dalje, $C \to B$ je netrivijalna funkcijska zavisnost, a njena lijeva strana ne sadrži niti jedan ključ za (R, F). Zbog toga, (R, F) nije u $BCNF$.

(b) Dekomponiranje R preko funkcijske zavisnosti $C \to B$ daje komponente $R_1 : CB$, $R_2 : CA$. Kako smo dobili komponente od samo dva atributa, zaključujemo temeljem Propozicije (nestandardna zavisnost i normalne forme) da je $d(R) : \underline{C}B, \underline{CA}$ $BCNF$ dekompozicija. Ključevi komponenti su podcrtani. Osim toga, kako je dekompozicija dobivena preko funkcijske zavisnosti, znamo da ona čuva informaciju. (Sjetimo se da $X \to Y \models \bowtie(XY, X(R - XY)))$.

Primjer

Primjer

(4NF)

$R = ABCD,\ F : A \to B, B \to C, B \twoheadrightarrow CD.$

(a) *ispitati je li* (R, F) *u* $4NF$

(b) *ako* (R, F) *nije u* $4NF$, *onda odredite* $4NF$ *dekompoziciju koja čuva informaciju.*

Rješenje I

Rješenje

(a) Propozicija (ključni trik) vrijedi i u slučaju da F sadrži funkcijske i višeznačne zavisnosti. Atributi, koji nisu u desnoj strani funkcijskih zavisnosti iz F, participiraju u ključu za (R, F). Prema tome $AD \subseteq K$. Dalje, $F \models AD \rightarrow R$. Zato je $K = AD$ jedini ključ za (R, F). Kako vrijedi $F \models A \twoheadrightarrow B$ i $A \twoheadrightarrow B$ je netrivijalna višeznačna zavisnost čija lijeva strana ne sadrži ključ za (R, F), zaključujemo da (R, F) nije u $4NF$.

Rješenje II

(b) Dekomponiranje skupa R preko višeznačne zavisnosti $A \twoheadrightarrow B$ daje komponente $R_1 : AB$, $R_2 : ACD$. Komponenta R_1 ima samo dva atributa, a budući F ne sadrži nestandardnu višeznačnu zavisnost (zavisnost oblika $\emptyset \twoheadrightarrow Y$), dobivamo da je R_1 u $4NF$. Preostaje ispitati komponentu $R_2 : ACD$. Ključ za $(R_2, \Pi_F(R_2))$ je $K = AD$. Naime, vrijedi da svaki ključ za (R, F) jeste i ključ za $(R_i, \Pi_F(R_i))$ pod uvjetom da je $K \subseteq R_i$ (ovo svojstvo se naziva nasljeđivanje ključa). Također, i svojstvo 'biti jedini ključ' se nasljeđuje. U nastavku našeg primjera, imamo da je $K = AD$ jedini ključ za $(R_2, \Pi_F(R_2))$. Budući vrijedi $\Pi_F(R_2)) \models A \twoheadrightarrow C$ i $A \twoheadrightarrow C$ je netrivijalna višeznačna zavisnost za $R_2 : ACD$ čija lijeva strana ne sadrži ključ $K = AD$, zaključujemo da $(R_2, \Pi_F(R_2))$ nije u $4NF$.

Rješenje III

Dekomponiranjem skupa $R_2 : ACD$ preko višeznačne zavisnosti $A \twoheadrightarrow C$ dobivamo komponente $R_{21} : AC$, $R_{22} : AD$. Konačno, $d(R) : AB, AC, AD$ je $4NF$ dekompozicija koja čuva informaciju. Da $d(R)$ čuva informaciju proizlazi iz postupka dekomponiranja preko višeznačne zavisnosti i činjenice da vrijedi $X \twoheadrightarrow Y \equiv \bowtie(XY, X(R - XY))$.

Primjer

Primjer

(5NF)

$R = AB,\ F : \bowtie(A, B)$

(a) *ispitati je li* (R, F) *u* $5NF$

(b) *ako* (R, F) *nije u* $5NF$*, onda odredite* $5NF$ *dekompoziciju koja čuva informaciju.*

Teorija baza podataka
Modeliranje i normalizacija baza podataka

Uvod
Zavisnosti u relacijskim bazama podataka
Formalni sustavi
Implikacijski problem
Normalne forme
Pitanja?

Rješenje

(a) Budući je $K = AB$ jedini ključ za (R, F) i $\bowtie(A, B)$ je netrivijalna zavisnosti spoja sa svojstvom da niti jedna njena komponenta ne sadrži ključ (dovoljno je da barem jedna ne sadrži ključ), zaključujemo da (R, F) nije u $5NF$.

(b) Dekomponiranjem skupa R preko netrivijalne zavisnosti spoja $\bowtie(A, B)$ dobivamo $5NF$ dekompoziciju $d(R) : A, B$, koja čuva informaciju.

Primjer

Primjer

(6NF)

$R = AB$, $F : A \to B$

(a) *ispitati je li* (R, F) *u* $6NF$

(b) *ako* (R, F) *nije u* $6NF$*, onda odredite* $6NF$ *dekompoziciju koja čuva informaciju.*

Rješenje

Rješenje

Budući da u F^+ ne postoji netrivijalna zavisnost spoja, zaključujemo da je (R, F) u $6NF$.

Zadaci

Zadatak

$R = ABCDE$, $F : AB \to C,\ C \to D,\ CD \to A,\ A \to BE$

Primijeniti algoritam sinteze 3NF.

Zadatak

$R = ABCD$, $F : A \to C,\ B \twoheadrightarrow A$

(a) *Ispitati je li* (R, F) *u* $4NF$

(b) *Ako* (R, F) *nije u* $4NF$*, onda odredite* $4NF$ *dekompoziciju koja čuva informaciju*

Zadaci

Zadatak

$R = ABCDE,\ F : AB \to C,\ \bowtie(AB, ABCDE)$

(a) *Je li* (R, F) *u* $6NF$?

(b) *Je li* (R, F) *u* $5NF$?

(c) *Je li* (R, F) *u* $4NF$?

(d) *Je li* (R, F) *u* $BCNF$?

(e) *Je li* (R, F) *u* $3NF$?

(f) *Je li* (R, F) *u* $2NF$?

Pitanja?

Izvor

Maleković, M., Schatten, M. (2017) Teorija i primjena baza podataka, Fakultet organizacije i informatike, Varaždin.