Dr. Mirko Maleković
Fakultet organizacije i informatike
Varaždin

UDK: 681.327.2
Izvorni znanstveni rad

Čuvanje funkcijskih zavisnosti u postupku dekomponiranja relacijske sheme baze podataka

Poželjno svojstvo dekompozicije relacijske sheme baze podataka je čuvanje zavisnosti. U ovom radu razmatramo čuvanje zavisnosti u postupku dekomponiranja relacijske sheme (R,F), gdje je F skup funkcijskih zavisnosti. Beeri i Honeyman predložili su algoritam za testiranje čuvanja funkcijskih zavisnosti (algoritam se bazira na R_i-operaciji, a opisan je u [Ullman 88]). Dokazali smo svojstva R_i-operacije, koja su omogućila modifikaciju navedenog algoritma eliminiranjem suvišnih R_i -operacija.

Ključne riječi: algoritam za testiranje čuvanja zavisnosti, dekompozicija relacijske sheme, funkcijska zavisnost, logička posljedica, R_i-operacija.

1 Uvod

Dobra shema relacijske baze podataka treba imati određena, poželjna svojstva; uz čuvanje informacije, uobičajeni zahtjev je i čuvanje zavisnosti ([Gallaire et al. 81], [Honeyman 82], [Maier 83], [Ullman 88] i [Vardi 88]).

U slučaju dekomponiranja relacijske sheme (R,F), gdje je F skup funkcijskih zavisnosti, Beeri i Honeyman su dali algoritam za testiranje čuvanja zavisnosti; algoritam je naveden u [Ullman 88].

U ovom radu iskazujemo određene činjenice vezane za čuvanje zavisnosti (formulacija čuvanja zavisnosti u slučaju skupa funkcijskih zavisnosti, i svojstva R_i-operacije), koje omogućavaju modifikaciju prije spomenutog algoritma.

Osim uvodnog dijela i dodatka članak se sastoji od 4 odjeljka. U odjeljku 2 dajemo osnovne pojmove relacijskog modela. Karakterizacija dekompozicije koja čuva funkcijske zavisnosti i Algoritam (Beeri + Honeyman) za testiranje čuvanja funkcijskih zavisnosti dani su u odjeljku 3. Svojstva R_i-operacije prezentirana su u odjeljku 4. Koristeći ta svojstva, modificirali smo Algoritam (Beeri+Honeyman) tako da smo eliminirali suvišne R_i-operacije.

Zaključak je dan u odjeljku 5. U dodatku A dan je dokaz Propozicije (FZ) koja govori o čuvanju funkcijskih zavisnosti. Dokaz Propozicije (svojstva R_i-operacije) naveden je u dodatku B.

2 Osnovni pojmovi

Relacijska shema je uređen par (R,F), gdje je R konačan, neprazan skup atributa, a F je skup zavisnosti nad R. Pretpostavljamo da je svakom atributu A iz R pridružen neprazan skup, u oznaci $dom(A_i)$, koji se zove domena atributa A_i.

Dalje, $D = \bigcup_{A_i \in R} dom(A_i)$ je domena od R.
Slog nad R je preslikavanje $t : R \to D$, tako da vrijedi $t(A_i) \in dom(A_i)$ za svaki $A \in R$.

Neka je t slog nad R i $Z \subseteq R$. Sa t[Z] označavamo projekciju sloga t na skup atributa Z.

Relacija nad R je uređen par (R,r) , gdje je r konačan skup slogova nad R.

Neka su $X, Y \subseteq R$. Funkcijska zavisnost je izraz oblika $X \to Y$. Kažemo da $X \to Y$ vrijedi u (R, r), u oznaci $(R, r)| \vdash X \to Y$, ako i samo ako

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

Jednakošću $FZ(R) = \{X \to Y \mid X, Y \subseteq R\}$ definiran je skup svih funkcijskih zavisnosti nad R. Neka je $F \subseteq FZ(R)$. Uređen par (R, F) zovemo funkcijski zavisna relacijska shema. Kako u daljem radimo samo s funkcijskim zavisnostima, (R, F) ćemo kratko zvati relacijska shema. Kažemo da skup funkcijskih zavisnosti F vrijedi u relaciji (R, r), u oznaci $(R, r)| \vdash F$, ako i samo ako

$$(\forall f \in F)[(R, r) \,|\vdash f].$$

Neka su $F, G \subseteq FZ(R)$. Kažemo da je G logička posljedica od F, u oznaci $F \models G$, ako i samo ako

$$(\forall (R, r))[(R, r) \,|\vdash F \Rightarrow (R, r)| \vdash G].$$

Ako je $G = \{g\}$, tada umjesto $F \models \{g\}$ pišemo kraće $F \models g$.

Neka je $F \subseteq FZ(R)$ i $X \subseteq R$. Jednakostima:
$F^+ = \{f \in FZ(R) \mid F \models f\}$ i $X_F^+ = \{A \in R \mid F \models X \to A\}$, definirani su zatvarač od F i zatvarač od X s obzirom na F, respektivno.

Neka je (R, F) relacijska shema. Familiju skupova $d(R) = \{R_1, .., R_k\}$, gdje je $R_1, .., R_k \subseteq R$, zovemo dekompozicijom od (R, F) ako i samo ako vrijedi:

(1) $R_i \neq \emptyset$, za svaki R_i iz $d(R)$, i

(2) $\bigcup_{i=1}^{k} R_i = R$.

Sada definiramo projekciju relacije (R, r) na skup atributa X, $X \subseteq R$, te projekciju skupa funkcijskih zavisnosti F na skup atributa $X, X \subseteq R$. Definicije su dane slijedećim jednakostima:

$\prod_X(R, r) = (X, s)$, gdje je $s = \{t[X] \mid t \in r\}$, je projekcija relacije (R, r) na skup atributa X;

$\prod_X(F) = \{Y \rightarrow Z \mid Y, Z \subseteq X \text{ i } F \models Y \rightarrow Z\}$ je projekcija skupa funkcijskih zavisnosti F na skup atributa X.

U dokazima, koji su dani u dodatku ovog rada, koristit ćemo slijedeća svojstva:

(NFZ) $\prod_X(R, r) \mid\vdash F \Rightarrow (R, r) \mid\vdash F$
(neugrađenost funkcijskih zavisnosti)

(DFZ) $F \models X \rightarrow Y \Rightarrow F \models X \rightarrow Z$ ako $Z \subseteq Y$
(dekompozicija za funkcijske zavisnosti)

(TFZ) $F \models X \rightarrow Y \wedge F \models Y \rightarrow Z \Rightarrow F \models X \rightarrow Z$
(tranzitivnost za funkcijske zavisnosti)

Dokazi navedenih svojstava vrlo su jednostavni i zbog toga ispušteni.

3 Dekompozicija koja čuva funkcijske zavisnosti

Za dekompoziciju $d(R, F) = \{R_1, .., R_k\}$ relacijske sheme (R, F) kažemo da čuva zavisnosti F ako i samo ako

$$(\forall(R, r))\, [\prod_{R_1}(R, r) \mid\vdash \prod_{R_1}(F) \wedge .. \wedge \prod_{R_1}(R, r) \mid\vdash \prod_{R_1}(F) \Rightarrow (R, r) \mid\vdash F]$$

Navedena definicija kaže slijedeće: Umjesto baze podataka, koja se sastoji od jedne relacije (R, r) ("velika" tablica), imamo bazu podataka koju čine relacije $\prod_{R_1}(R, r), .., \prod_{R_i}(R, r)$ ("male" tablice). Relacija (R, r) je valjana ako $(R, r) \mid\vdash F$ (globalna valjanost). Analogno, relacije $\prod_{R_1}(R, r), .., \prod_{R_1}(R, r)$ su valjane ako $\prod_{R_1}(R, r) \mid\vdash \prod_{R_1}(F) \wedge .. \wedge \prod_{R_i}(R, r) \mid\vdash \prod_{R_i}(F)$ (lokalne valjanosti).

U definiciji čuvanja zavisnosti zahtijeva se da lokalne valjanosti povlače globalnu valjanost.

Vrijedi slijedeća propozicija.

Propozicija (FZ)

Neka je $d(R, F) = \{R_1, .., R_k\}$ dekompozicija relacijske sheme (R, F), gdje je F skup funkcijskih zavisnosti. Tada,

$$d(R, F)\,\text{čuva}\,F \Longleftrightarrow \prod_{R_1}(F) \cup .. \cup \prod_{R_i}(F) \models F$$

Dokaz Propozicije (FZ) dan je u dodatku A.

Prema Propoziciji(FZ), čuvanje funkcijskih zavisnosti reducira se na testiranje implikacijskog problema

$$\text{(IP)} \quad \prod_{R_1}(F) \cup .. \cup \prod_{R_i}(F) \models F$$

Osnovna ideja Algoritma (Beeri+Honeyman) je testiranje (IP) bez računanja $\prod_{R_1}(F) \cup .. \cup \prod_{R_i}(F)$ (navedeno računanje je eksponencijalne kompleksnosti, jer treba računati F^+). Uveli su R_i-operaciju, koja se definira jednakošću:

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

Koristeći R_i-operacije, testiranje implikacijskog problema $G \models X \rightarrow Y$, gdje je $G = \prod_{R_1}(F) \cup .. \cup \prod_{R_i}(F)$, a $X \rightarrow Y \in F$, vrši se pomoću algoritma:

Algoritam (Beeri + Honeyman)

Ulaz: (R, F), $d(R, F) = \{R_1, .., R_k\}$, $X \to Y \in F$

Izlaz: DA ako $G \models X \to Y$
NE ako ne vrijedi $G \models X \to Y$

Postupak:
(1) Računati X_G^+ pomoću modula $M(X_G^+)$

(2) $G \models X \to Y$ ako $Y \subseteq X^+$
$G \models X \to Y$ ne vrijedi ako ne vrijedi $Y \subseteq X_G^+$

<u>Modul (X_G^+)</u>

```
begin
  Z := X
  While Z se mijenja do
    for j = 1 to k do
      Z := Z ∪ [(Z ∩ R_j)^+_F ∩ R_j]
  endwhile
  X^+_G = Z
end
```

Konačno, testiranje $G \models F$ reducira se na testiranje $G \models f$ za svaku zavisnost $f \in F$.

Pri tome,

$G \models F$ ako $(\forall f \in F)[G \models f]$,
$G \models F$ ne vrijedi ako $(\exists f \in F)[G \models f$ ne vrijedi $]$.

Centralna stvar u testiranju $G \models F$ je modul $M(X_G^+)$. Modifikaciju modula $M(X_G^+)$ ostvarit ćemo koristeći svojstva R_i -operacije ; modifikacija se sastoji u eliminaciji suvišnih R_i -operacija.

4 Svojstva R_i -operacije

R_i operacija ima slijedeća svojstva:

Propozicija (svojstva R_i-operacije)

Neka je (R, F) relacijska shema i $X, R \subseteq R$. Tada,

(A) $R_i \subseteq X \Rightarrow R_i(X, F) = X$

(B) $X \cap R_i = \emptyset \Rightarrow R_i(X, F) = X$ ako F ne sadrži zavisnosti oblika $\emptyset \rightarrow V$, gdje je $V \subseteq R$ i $V \neq \emptyset$.

(C) $R_i[R_i(X, F)] = R_i(X, F)$

Dokaz Propozicije (svojstva R_i-operacije) dan je u dodatku B.
Svojstvo (A) kaže da R_i-operacija ne mijenja X ako je $R \subseteq X$.
Svojstvo (B) iskazuje da R_i-operacija ne mijenja X ako F ne sadrži zavisnosti oblika $\emptyset \rightarrow V, V \subseteq R$ i ako je $X \cap R = \emptyset$.
Svojstvo (C) kaže da R_i-operacija ne mijenja skup $R_i(X, F)$.
Primjenom Propozicije (svojstva R_i-operacije) možemo modificirati Algoritam (Beeri + Honeyman) tako da isključimo one R_i-operacije koje, na osnovi svojstava (A), (B) i (C), ne mijenjaju skup Z (iz modula $M(X_G^+)$) Modificirani modul $M(X_G^+)$, u oznaci $MODIFM(X_G^+)$, ima oblik:

MODIFM (X_G^+)

Ulaz: (R, F), $d(R, F) = \{R_1, .., R_k\}$, $X \subseteq R$

Izlaz: X_G^+, gdje je $G = \prod_{R_1}(F) \cup .. \cup \prod_{R_i}(F)$.

Postupak:

```
begin
  Z := X
   while Z se mijenja do
    if F ne sadrži zavisnosti oblika ∅ → V,  V ≠ ∅ then

          for j := 1 to k takav da
            [R_j ⊈ Z ∧ R_j ∩ Z ≠ ∅ ∧ R_j-operacija nije posljednja
             koja je promijenila Z] do
            Z := Z ∪ [(Z ∩ R_j)^+_F ∩ R_j]

          else

          for j := 1 to k takav da
            [R_j ⊈ Z∧ R_j-operacija nije posljednja koja je
             promijenila Z] do

            Z := Z ∪ [(Z ∩ R_j)^+_F ∩ R_j
        endwhile

  X_G^+ := Z
end
```

Propozicija (korektnost modula MODIFM (X_G^+))

Modul MODIFM (X_G^+) je korektan.

Dokaz

Modul MODIFM (X_G^+) sastoji se od dvije for-petlje. Prva for-petlja je primijenjena ako F ne sadrži zavisnosti oblika $\emptyset \rightarrow V$, gdje je $V \subseteq R$ i $V \neq 0$. U tom slučaju, suvišne su one R_j-operacije, gdje je $R \subseteq Z$ ili $R \cap Z = \emptyset$ ili R_j-operacija je posljednja koja je promijenila Z (redom su primijenjena svojstva (A), (B) i (C) Propozicije (svojstva R_i-operacije)). Prema tome, treba primijeniti one R_j-operacije za koje vrijedi $R_j \not\subseteq Z$ i $R_j \cap Z = \emptyset$ i R_j-operacija nije posljednja koja je promijenila Z. Druga for-petlja je primijenjena ako F sadrži zavisnosti oblika $\emptyset \rightarrow V$, gdje je $V \subseteq R$ i $V \neq \emptyset$. U tom slučaju, R_j-operacija, za koju je $R_j \cap Z = \emptyset$, može mijenjati Z te ju treba izvršiti. Zato, druga for-petlja isključuje one R_j-operacije za koje vrijedi $R_j \subseteq Z$ ili R_j-operacija je posljednja koja je promijenila Z . Prema tome, u drugoj for-petlji primijenjene su one R_j-operacije za koje je $R_j \not\subseteq Z$ i R_j-operacija nije posljednja koja je promijenila Z. Kako se for-petlje modula MODIFM (X_G^+) ponašaju korektno i kako je modul Modul (X_G^+) iz Algoritma (Beeri+Honeyman) korektan (vidjeti dokaz u [Ullman 88]), zaključujemo da je modul MODIFM (X_G^+) korektan.

5 Zaključak

U ovom radu razmatrali smo čuvanje funkcijskih zavisnosti u postupku dekomponiranja relacijske sheme. Algoritam (Beeri + Honeyman), koji testira čuvanje funkcijskih zavisnosti, modificiran je na osnovi propozicije o svojstvima R_i-operacije. Modifikacija je ostvarena tako da su eliminirane suvišne R_i-operacije u postupku računanja X_G^+, gdje je $X \subseteq R$, $G = \prod_{R_1}(F) \cup .. \cup \prod_{R_k}(F)$, $F \subseteq FZ(R)$ i $d(R) = \{R_1, .., R_k\}$. Implementacijom dobivenog algoritma može se dobiti koristan alat za testiranje da li shema relacijske baze podataka ima svojstvo čuvanja funkcijskih zavisnosti.

DODATAK A. DOKAZ PROPOZICIJE (FZ) U ovom dodatku dokazujemo Propoziciju (FZ) iz odjeljka 3. Propozicija utvrđuje da dekompozocija $d(R) = \{R1, .., R_k\}$ relacijske sheme (R, F), gdje je F skup funkcijskih zavisnosti, čuva F ako i samo ako

$$\bigcup_{i=1}^{k} \prod_{R_i}(F) \models F$$

Dokazujemo smjer ($\Rightarrow$).

Neka d(R) čuva F. Trebamo dokazati da vrijedi $\bigcup_{i=1}^{k} \prod_{R_i}(F) \models F$.

Uzmimo proizvoljnu relaciju (R, r) nad R takvu da vrijedi $(R, r) \mid\vdash \bigcup_{i=1}^{k} \prod_{R_i}(F)$. Pokazat ćemo da tada vrijedi $(R, r) \models F$.

Jer $(R, r) \mid\vdash \bigcup_{i=1}^{k} \prod_{R_i}(F)$, dobivamo $(R, r) \mid\vdash \prod_{R_i}(F)$ za svaki R_i.
Zato, $\prod_{R_i}(R, r) \mid\vdash \prod_{R_i}(F)$, za svaki R_i. Kako $d(R)$ čuva F , slijedi $(R, r) \mid\vdash F$.

Dokazujemo smjer ($\Leftarrow$).

Neka vrijedi $\bigcup_{i=1}^{k} \prod_{R_i}(F) \models F$. Dokazat ćemo da tada $d(R)$ čuva F. Pretpostavimo da vrijedi

$$\prod_{R_1}(R, r) \mid\vdash \prod_{R_1}(F) \wedge .. \wedge \prod_{R_k}(R, r) \mid\vdash \prod_{R_k}(F).$$

Trebamo dokazati $(R, r) \mid\vdash F$. Iz pretpostavke da $\prod_{R_i}(R, r) \mid\vdash \prod_{R_i}(F)$ vrijedi za svaki R_i, dobimo $(R, r) \mid\vdash \prod_{R_i}(F)$, za svaki R_i (neugrađenost funkcijskih zav.). Zato, $(R, r) \mid\vdash \bigcup_{i=1}^{k} \prod_{R_i}(F)$. Kako smo pretpostavili

$$\bigcup_{i=1}^{k} \prod_{R_i}(F) \models F, \text{ dobivamo } (R, r) \mid\vdash F.$$

DODATAK B . DOKAZ PROPOZICIJE (svojstva R_i-operacije)
U ovom dodatku dokazujemo Propoziciju (svojstva R_i-operacije), koja je navedena u odjeljku 4. Propozicija utvrđuje:
Neka je (R, F) relacijska shema i $X, R \subseteq R$.
Tada,

(A) $R \subseteq X \Rightarrow R_i(X, F) = X$

(B) $X \cap R_i = \emptyset \Rightarrow R_i(X, F) = X$ ako F ne sadrži zavisnosti oblika $\emptyset \rightarrow V$, gdje je $V \subseteq R$ i $V \neq \emptyset$.

(C) $R_i[R_i(X, F)] = R_i(X, F)$

Dokaz svojstva (A).

Iz pretpostavke $R_i \subseteq X$, direktno slijedi

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

Dokaz svojstva (B).

Neka je (R, F) relacijska shema, gdje F ne sadrži funkcijske zavisnosti oblika $\emptyset \rightarrow V$, $V \subseteq R$, $V \neq \emptyset$. Pretpostavimo, takodjer, $X \cap R_i = \emptyset$. Trebamo dokazati $R_i(X, F) = X$.
Iz $X \cap R_i = \emptyset$ slijedi $(X \cap R_i)^+_F = \emptyset$ jer smo pretpostavili da F ne sadrži zavisnosti oblika $\emptyset \rightarrow V$, $V \subseteq R$, $V \neq \emptyset$. Zato, $R_i(X, F) = X \cup [(X \cap R_i)^+_F \cap R_i] = X \cup [\emptyset \cap R_i] = X \cup \emptyset = X$.

Dokaz svojstva (C).

Trebamo dokazati $R_i[R_i(X, F)] = R_i(X, F)$.
Iz $R_i[R_i(X, F)] = R_i(X, F) \cup [(R_i(X, F) \cap R_i)^+_F \cap R_i]$ vidimo da vrijedi $R_i(X, F) \subseteq R_i[R_i(X, F)]$. Preostaje dokazati da vrijedi $R_i[R_i(X, F)] \subseteq R_i(X, F)$. Neka je $A \in R_i[R_i(X, F)]$. Dokazat ćemo da je tada $A \in R_i(X, F)$. Iz $A \in R_i[R_i(X, F)]$ slijedi $A \in R_i(X, F)$ ili $A \in [(R_i(X, F) \cap R_i)^+_F \cap R_i]$. Ako je $A \in R_i(X, F)$, onda je to upravo ono što trebamo dokazati. Zato, pretpostavimo $A \in [(R_i(X, F) \cap R_i)^+_F \cap R_i]$. Dokazat ćemo da iz navedene pretpostavke slijedi $A \in R(X, F)$. Iz $A \in [(R_i(X, F) \cap R_i)^+_F \cap R_i]$ slijedi $A \in (R_i(X, F) \cap R)^+_F$ i $A \in R_i$. No, $A \in (R(X, F) \cap R_i)^+_F$ znači da vrijedi
(I) $F \models (R_i(X, F) \cap R_i) \rightarrow A$. Dalje, $F \models (X \cap R_i) \rightarrow (X \cap R_i)^+_F$ (iz definicije zatvarača skupa atributa s obzirom na skup zavisnosti F) . Po pravilu dekompozicije za funkcijske zavisnosti, dobivamo:
(II) $F \models (X \cap R_i) \rightarrow (X \cap R_i)^+_F \cap R_i$.

Računajući $R_i(X, F) \cap R_i$, dobit ćemo:

$$
\begin{aligned}
R_i(X,F)\cap R_i = {} & [X\cup[(X\cap R_i)^+_F\cap R_i]]\cap R_i = (X\cap R_i)\cup \\
& [((X\cap R_i)^+_F\cap R_i)\cap R_i] = \\
& (X\cap R_i)\cup[(X\cap R_i)^+_F\cap R_i] = \\
& [(X\cap R_i)\cup(X\cap R_i)^+_F]\cap[(X\cap R_i)\cup R_i] = \\
& (X\cap R)^+_F\cap R_i.
\end{aligned}
$$

Posljednja jednakost je posljedica činjenice da vrijedi
$(X\cap R_i)\subseteq R_i$ i $(X\cap R_i)\subseteq(X\cap R_i)^+_F$. Dakle, dobili smo
(III) $R_i(X,F)\cap R_i=(X\cap R_i)^+_F\cap R_i$.
Iz (II) i (III) dobivamo:
(IV) $F\models(X\cap R_i)\rightarrow(R_i(X,F)\cap R_i)$.
Konačno, iz (IV) i (I) , primjenom tranzitivnosti za funkcijske zavisnosti, dobivamo
$F\models(X\cap R_i)\rightarrow A$ tj. $A\in(X\cap R_i)^+_F$, što je i trebalo dokazati.

Literatura

[Gallaire et al. 81] H. Gallaire, J. Minker, and J.M. Nicolas, Advances in Database Theory, Vol. I, Plenum Press, New York, (1981).

[Honeyman 82] P. Honeyman, Testing satisfaction of functional dependencies, J. ACM 29:3 (1982), pp. 668-677.

[Maier 83] D. Maier, The theory of Relational Databases, Computer Science Press, Rockville, Md., (1988).

[Ullman 88] J.D. Ullman, Principles of Database and Knowledge Base Systems, Volume I, Computer Science Press, (1988).

[Vardi 88] M.Y. Vardi, Fundamentals of dependency theory, in Trends in Theoretical Computer Science (E. Borger, ed.), Computer Science Press, Rockville, Md., (1988), pp. 171-224.

Primljeno: 1993-06-17

Maleković M. Preservation of functional dependencies in decomposing relational database scheme

SUMMARY

It is desirable for a decomposition of relational scheme to have the dependency-preserving property. In this work, the dependency-preserving is considered for the decomposition of relational scheme (R, F), where F is a set of functional dependencies. A modification of Beeri and Honeyman's algorithm for testing the preservation of functional dependencies (the algorithm is described in [Ullman 88]) is presented. The modification is based on the properties of R_i operation.